THE
AUSTRALIAN
NATIONAL
UNIVERSITY

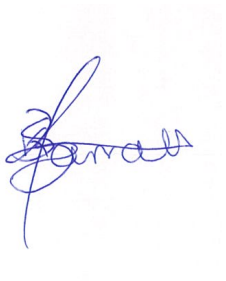
New aspects of *ZIC2*-associated Holoprosencephaly

Kristen S. Barratt

A thesis submitted for the degree of Doctor of Philosophy of the Australian National University

December 2017

Except where otherwise acknowledged, the work presented in this thesis is my own.



Kristen Barratt

Work completed by others in this thesis:

Chapter 5: The optimization of the RT-qPCR protocol, cloning and mRNA decay analysis were developed and performed in conjunction with Kathryn Dickson, an Honours student in the Arkell lab.

Chapter 7: The work on *PiggyBac* transgenics was performed in conjunction with the Muenke laboratory at the National Human Genome Research Institute (NHGRI) and the Transgenic Core Facility (NIH, Maryland, USA).

The work on CRISPR mutagenesis was performed in conjunction with the Burgio laboratory at the John Curtin School of Medical Research (JCSMR) Transgenesis Facility (ANU, Canberra, Australia). These facilities aided in the production of transgenic and CRISPR mutants

Publications from work in this thesis:

Chapter 1: Barratt, K.S. & Arkell, R.M (2017) Chapter 14: *ZIC2* in Holoprosencephaly, "*Zic* family - Evolution, Development and Disease", Springer (in publication).

Chapter 3: Barratt, K.S., Glanville-Jones, H.C. & Arkell, R.M., 2014. The *Zic2* gene directs the formation and function of node cilia to control cardiac situs. Genesis, 52(6), pp.626–635.

Diamand, K.E.M., Barratt, K.S. & Arkell, R.M (2017) Chapter 10: Overview of rodent *Zic* genes, "Zic family - Evolution, Development and Disease", Springer (in publication).

Appendix: Barratt, K.S., Diamand, K. E. M. & Arkell, R.M, Validation of Reference Genes for RT-qPCR Studies of Murine Gastrulation, Manuscript awaiting submission, 2017.

Acknowledgements

This PhD has taken me a long time, so there's a large list of people to thank.

First and foremost, I'd like to thank Ruth Arkell. My time as a PhD student in your lab has afforded me fantastic opportunities to develop my skill set as a researcher and as a teacher. You have mentored me and made sure that I did not give up on this project when everything was going wrong. You also allowed me to go at my own pace and tried your best to leave me alone, which I'm sure was frustrating. Thank you for your supervision.

I also want to acknowledge essential members of the Arkell lab, past and present, who have provided me with emotional, psychological and technical support for my entire PhD – Helen Bellchambers, Koula Diamand, Hannah Glanville-Jones, Kathryn Dickson, Jerry Ahmed, Radiya Ali, Nicole Thompson, Adelaide Dennis and Alaa Alzahrani. Without your knowledge, your ability to listen to me rant, your terrible jokes and your endless chocolate supply, I would not have made it through this degree.

To anyone who made me laugh or smile when I really needed it, gave me a job when I needed it, listened to me vent my frustration when things had stalled, had to endure my fun facts about lobsters, whales and space mice, or indulged me in stupid arguments – thank you. You will never know how much I needed and appreciated your kindness, friendship, and terrible puns. I'd especially like to thank all my besties over at the teaching labs, Adelaide (without whom this thesis would not have been written), Kathryn and Lora, and my adopted family of Sarah, Andrew, Janelle and Emily. You may all be crazy, but you're my kind of crazy.

I'd particularly like to thank my partners in crime Koula, Hannah and Helen (aka The Oracle). I'd never admit this to your face, but I kind of like you guys. You've all helped me, taught me and laughed with me. You listened to my terrible jokes (mostly) without complaint and my whining (mostly) without judgment. You are the sounding board for my terrible ideas, and the skills and techniques that you have taught me are priceless and the main reason this thesis has any results. Thank you for allowing me to bug you with the most mundane questions, even on the days when you really didn't have the time or the energy.

Hannah – we've been side by side our entire PhD's (with some detours on the way), and you have been my cheerleader and work-spouse the whole way. I hope that I've returned the favour somewhat. You are also the closest thing to an adult in this lab, which has proven essential an innumerable number of times.

Koula – the game is pronounced eww-no not you-know, drizzle is lighter rain than sprinkling, and fun fact *are* cool. But also, thank you for the delicious lamb, stupid conversations and your

company while I scream into the void. Since you demanded a poem dedicated to you, I wrote you a haiku because I'm no good at odes:

Your friendship is like

A donut on a leafy

Green. Not a salad.

The Muenke lab at the NHGRI, NIH and the transgenic team at the NIH, especially Erich Roessler and Gene Elliot, welcomed me to the NIH with open arms for three months of my PhD. They were willing to let me use their money and resources to test out a new technique that only kind-of worked. My experience visiting their lab and learning new techniques was priceless and I hope that I get the opportunity to work with them again. Also, thanks for all the delicious cheesecake.

I'd also like to thank Gaetan Burgio, Nay Chi and everyone else at the JCSMR Transgenesis Facility for their hard work. Despite the fact that the *Zic2* 3'UTR never does anything you want it to do, they still haven't given up on making the CRISPR mutants for us. What idiots.

I must also thank chocolate and caffeine, which have sponsored me through this degree, as well as the multiple different musicians and podcasts that have entertained me throughout. If I have any sanity left, it's because of them. I'd also like to thank the Australian Government for their support through the Australian Postgraduate Award.

Finally, I am eternally indebted to my family for their support during my PhD, both those human family members and those members with four paws. Mum and Dad - you were always ready with a hug when I felt like I wasn't going to make it through. You gave me a sense of adventure and inability to give up, which is the reason I got myself into this mess in the first place. I'm also thankful for Ellise and Rob. It may not always seem like it, but you guys helped me through.

To Ivy – what's a crazy cat lady without her cat? Thank you for always purring when I come home at the end of a long day.

To my body, which has tried to kill me more than once during my PhD – you can stop now. We made it. It's time to relax.

I'd also like to give a special shout out to the *ZIC2* 3'UTR. You are my nemesis. You may have won this round, but I'm not ready to give up on you yet.

Thank you to you all and I hope we never have to do this again.

Kristen

For Milo, Ivy and Mary Dunn

Because you never wanted to know how the PhD was going

Abstract

Holoprosencephaly (HPE) occurs due to incomplete division of the developing forebrain along the embryonic midline, resulting in a failure to form two distinct cerebral hemispheres. Affecting 1/250 human conceptuses, HPE is a leading cause of pre- and post-natal morbidity and mortality. Currently, pathogenic mutations in the coding region of fifteen genes have been implicated in both classic and middle interhemispheric variants of HPE. Mutation of *ZIC2* accounts for 9% of solved cases, making it the second most common causative HPE gene after SHH. Nevertheless, multiple aspects of *ZIC2*-associated HPE remain unexplored, including the mechanisms that underlie HPE-associated co-morbidities, and how *Zic2* expression is regulated in the gastrulating embryo. The work presented in this thesis uses in silico analysis, cultured human cells and mouse models to investigate these aspects of *ZIC2*-associated HPE.

Numerous *Zic2*-associated HPE probands exhibit cardiac anomalies, yet these defects are often viewed as secondary to the HPE phenotype and their relationship to ZIC2 function has not been investigated. Characterisation of the cardiac defects (that occur alongside HPE in a mouse model harbouring the *Zic2* severe loss-of-function *Kumba* allele) shows they arise due to a loss of asymmetric gene expression at the early-somite node and in the left lateral plate mesoderm. Furthermore, ZIC2 acts upstream of, and is required for, the correct formation and function of cilia in the mid-gastrula node. This is the same region of the murine embryo in which ZIC2 is required during normal development to prevent HPE, suggesting a common tissue of origin for the observed brain and cardiac defects, and that *ZIC2* mutation is a risk factor for the development of left-right defects.

Analysis by human geneticists identified single nucleotide variants within the *ZIC2* 3'UTR of otherwise unsolved HPE probands, potentially pinpointing a genomic region essential for the control of *ZIC2* expression during gastrulation. Characterisation of the *ZIC2* 3'UTR in a signalling environment reminiscent of the gastrula node indicates it contains a regulatory element that functions as a transcriptional repressor, as well as multiple transcript stability elements that regulate *ZIC2* half-life. This element warrants further in vivo assessment of the mechanism by which it controls *ZIC2* expression and evaluation of the pathogenicity of the known SNVs.

Table of Contents

List of Figures

List of Tables

Abbreviations

5’ cap	5’ 7-methylguanosine cap
6-FAM	6-carboxyfluorescein
ADA	Allelic Discrimination Assay
ADE	Anterior definitive endoderm
AER	Apical ectodermal ridge
ANC	Anterior notochord
ANOVA	Analysis of Variance
ANR	Anterior neural ridge
ANU	The Australian National University
AME	Anterior mesendoderm
A-P	Anterior-posterior
APS	Anterior primitive streak
ARE	AU-Rich Element
AREBPs	AU-Rich Element Binding Proteins
Argo	Argonaut
AUF1	AU-Rich Element RNA Binding Protein 1
AVE	Anterior visceral endoderm
BAC	Bacterial artificial chromosomes
β-gl	Beta globin minimal promoter
BMP	Bone Morphogenetic Protein
bp	Base pair
BRF1	Butyrate Response Factor 1
BSA	Bovine Serum Albumin
Cas	CRISPR Associated Protein
CBP	CREB Binding Protein
CDS	Coding DNA Sequence
CEF1	CUG-binding protein 1
Chrd	*Chordin*
cHS4	Chicken insulator fragments
CMV	Cytomegalovirus
Co-SMAD	Common mediator SMAD
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR RNA
C_T	Threshold Cycle

*Dand**	Dan domain BMP antagonist family member
DE	Definitive endoderm
DHS	DNase I Hypersensitivity Site
DIC	Differential Interference Contrast
*Dkk**	Dickkopf
*Dll**	Delta-like
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl Sulphoxide
dpc	Days post coitum
DRB	5,6-dichloro-beta D-ribofuranosyl-benzimidazole
DSB	Double stranded break
DTT	Dithiothreitol
D-V	Dorso-ventral
DVE	Distal visceral endoderm
E. coli	*Escherichia coli*
ED	Endoribunucleolytic Decay
EDTA	Ethylenediaminetetraacetic acid
ENU	*N*-ethyl-*N*-nitrosourea
ER element	Enhancer/Repressor element
EtOH	Ethanol
F12:DMEM	Ham's Nutrient Mixture F12: Dulbecco's Modified Eagle Medium
FBS	Foetal Bovine Serum
FGF	Fibroblast Growth Factor
FGFR	FGF tyrosine kinase receptors
FOX	Forkhead homeobox transcription factor
GHR	Growth hormone receptor
GM-CSF	Granulocyte Monocyte-colony stimulating factor
GRE	GU-rich Element
H3K4Me1	Monomethylation of histone 3 at lysine residue 4
H3K4Me3	Trimethylation of histone 3 at lysine residue 4
H3K27Ac	Acetylation of histone 3 at lysine residue 27
HDR	Homology-directed repair
HEK293T	Human Embryonic Kidney cells
hP0-3	Human poly(A) sites hP0, hP1, hP2 and hP3
HPE	Holoprosencephaly

HRP	Horse Radish Peroxidase
HRMA	High Resolution Melt Assay
Hsp68	Heat shock protein 68
IFT	Intraflagellar transport
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITR	Inverted terminal repeats
KSRP	KH-Type Splicing Regulatory Protein
Ku	*Kumba*
LB	Luria-Bertani
Lhx	LIM Homeobox
LLC-PK1	A ciliated kidney proximal tubule pig cell line
lncRNA	Long non-coding RNA
LPM	Lateral plate mesoderm
L-R	Left-right
LSD	Least Significant Difference
Luc	Luciferase Reporter
M1-6	Human HPE Mutation 1 to Mutation 6
MCS	Multiple cloning siite
MeOH	Methanol
MIHV	Middle interhemispheric variant
miRNA	MicroRNA
MM	Max Muenke; zebrafish GFP founder line designation
mNet	Mouse Notochord Enhancer of Transcription
mP0-3	Mouse poly(A) sites mP0, mP1, mP2 and mP3
mRNA	Messenger RNA
MW	Molecular Weight
ncRNAs	Non-coding RNAs
NCE	Non-coding Conserved Elements
NEB	New England Biolabs
NHEJ	Non-homologous end joining
NHGRI	National Human Genetics Research Institute
NIH	National Institute of Health
Nog	*Noggin*
NOTO	NOT Homeobox transcription factor
NPG	n-propyl gallate

opa	*Odd-paired*
ORF	Open reading frame
PABP	Poly(A) Binding Protein
PAM	Protospacer Adjacent Motif
PAP	Poly(A) Polymerase
PARN	Poly(A)-specific Ribonuclease
PAS	Poly(A) site
PBT	Phosphate Buffered Saline with Tween
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PFA	Paraformaldehyde
PK	Proteinase K
PKD	Polycystic kidney disease
Poly(A)	Polyadenylation
PrCP	Prechordal Plate
PS	Primitive streak
PTCH1	Patched
PVDF	Polyvinylidene Difluoride membrane
RACE	Rapid Amplification of cDNA Ends
RBP	RNA Binding Protein
RFX	Regulatory Factor X transcription factor
RISC	RNA-induced silencing complex
R-SMADs	Receptor-associated SMADs
RVNM	Rostral-ventral neural midline
RT-qPCR	Reverse transcription quantitative PCR
SACC	Subclass A C-terminally conserved domain
SANC	Subclass A N-terminally conserved domain
S. D.	Standard Deviation
SDS-PAGE	Sodium Dodecyl Sulphate – polyacrylamide gel electrophoresis
SEM	Scanning Electron Microscopy
S. E. M.	Standard Error of the Mean
sgRNA	Single-guide RNA
SHH	Sonic hedgehog
siRNA	Small interfering RNA
SMO	Smoothened

SNV	Single Nucleotide Variant
SOC	Super Optimal broth with Catabolite repression
TBE	Tris-Borate-EDTA Buffer
TBP	TATA-binding protein
Tbx	T-box transcription factor
TD	Touchdown PCR
TF	Transcription Factor
TFBS	Transcription Factor Binding Site
TIC	Translation Initiation Complex
T_m	Melting Temperature
tracrRNA	Trans-acting CRISPR RNA
TTP	Tristetraproline
U	Enzyme units
UCR	Ultraconserved Region
UTR	Untranslated Region
UV	Ultra-violet
VE	Visceral endoderm
WMISH	Whole Mount In Situ Hybridization
WNT	Wingless-related
WT	Wildtype
X-Gal	5-bromo-4-chloro-3-indolyl β-D-galactopyranoside
YAC	Yeast artificial chromosomes
ZFD	Zinc finger domain
ZF-NC	Zinc finger N-terminally conserved domain
ZIC	Zinc Finger of the Cerebellum transcription factor
ZOC	ZIC/Odd paired conserved motif
ZPA	Zone of polarising activity

Chapter 1: Introduction

This chapter contains text from the review: Barratt, K.S. & Arkell, R.M (2017) Chapter 14: *ZIC2* in Holoprosencephaly, "Zic family - Evolution, Development and Disease", Springer (in publication).

1.1 Holoprosencephaly

The evolution of bilaterans ~600 million years ago paved the way for the establishment of a vertebrate brain in which two hemispheres, divided along the embryonic midline, develop from a single group of cells. The embryological process that leads to hemisphere separation is estimated to completely or partially fail in approximately 1/250 human conceptuses, resulting in the most common structural defect of the human forebrain; Holoprosencephaly (HPE) (Matsunaga and Shiota, 1977). The crucial nature of hemisphere separation is evidenced by the positive correlation between the degree of brain malformation and HPE-induced mortality (Solomon et al. 2010b) and the reduction in HPE frequency to 1/10 000 by birth (Orioli and Castilla, 2010). The high rate of HPE occurrence suggests that the hemisphere separation process is incredibly fragile, comprised of multiple interconnected steps and vulnerable to interference. Furthermore, the phenotypic heterogeneity characteristic of this condition suggests that interference can come from a variety of sources (i.e. environmental and/or genetic factors) and that multiple factors act co-operatively in at least some HPE cases. The degree of brain separation and whether the failure occurs ventrally or dorsally distinguishes two main classes of HPE: classic and middle interhemispheric variant (MIHV).

1.1.1 Classic HPE

In classic HPE, the lack of separation is most severe ventrally, extending to the rostral, dorsal and posterior domains of the forebrain in a graded fashion. This leads to a spectrum of classic HPE (Table 1.1, Figure 1.1), of which alobar HPE (a monoventricle with no hemispheric separation) is the most severe form, followed by semilobar (partial hemispheric separation, resulting in fused left and right frontal and parietal lobes but retaining the posterior portion of the interhemispheric fissure) and lobar HPE (hemispheric and lateral vesicle separation is retained, except for in the rostral and ventral frontal lobes) (reviewed in Marcorelles and Laquerriere 2010). It is estimated that 10-40%, 43-45% and 17-33% of HPE cases are alobar,

Table 1.1: Human Holoprosencephaly types by severity*.

HPE Subdivisions	Anatomic Classification	Face and Brain Phenotypes generally observed
Alobar	Small single forebrain No interhemispheric division Asbsence of olfactory bulbs and tracts Absence or the corpus callosum Non separation of deep grey nuclei	Cyclopia (single or double eye) Proboscis, with or without a single nostril Distinct sockets Extreme ocular hypotelorism/ocuar hypotelorism (ie. close set eyes)
Semilobar	Rudimentary cerebral lobes Incomplete interhemispheric division Absence or hypoplasia of olfactory bults and tracts Absence of the corpus callosum Varying non separation of deep grey nuclei	Ocular hypotelorism Midline cleft lip Flat nose
Lobar	Fully developed cerebral lobes Dinstinct interhemispheric division Midline contnuous frontal neocortex Absent, hypoplasic or normal corpus callosum Separation of deep grey nuclei	Ocular hypotelorism Midline cleft lip (complete or partial) Single central maxillary incisor Flat nose Iris colombolar
Microform	No anomalies seen with conventional neuroimaging	Mild microcephaly Ocular hypotelorism Single maxillary central incisors
MIHV	Failure of separation of the posterior frontal and parietal lobes Callosal genu and splenium normally formed Absence of corpus callosum Hypothalamus and lentiform nuclei normally separated Heteotopic grey matter	Usually normal Hypertelorism sometimes observed

*Modified from Dubourg et al. 2007; Solomon, Lacbawan, et al. 2010 and Krauss 2007.

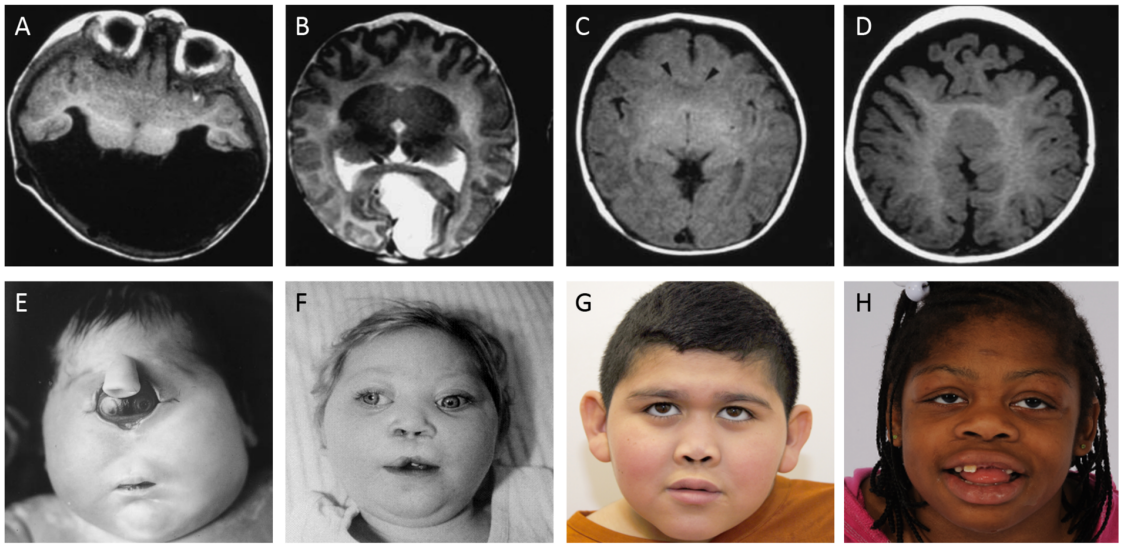


Figure 1.1: A comparison of the four subclasses of Holoprosencephaly. (a-d) Brain MRIs of human probands with (a) alobar, (b) semilobar, (c) lobar and (d) MIHV holoprosencephaly demonstrating the degree of separation between left and right hemispheres. **(e-h)** Facial phenotypes of human probands with (e) alobar, (f) semilobar, (g) lobar and (h) MIHV holoprosencephaly displaying a range of classical symptoms such as hypotelorism, cleft palates and a flattened nasal bridge. Figure is modified from (Solomon *et al.*, 2010c).

semilobar and lobar, respectively (Solomon *et al.*, 2010a); however, the number of cases at the severe end of the spectrum is likely is under-estimated due to early embryonic lethality.

In addition to hemisphere separation, HPE pathogenesis also comprises craniofacial and midline defects as common co-morbidities. The frequent coincidence of, and correlation between, the severity of brain and craniofacial symptoms in HPE is the basis for the notion that 'the face predicts the brain', as observed by DeMeyer and colleagues in 1964 (DeMyer *et al.*, 1964). Craniofacial phenotypes accompanying severe HPE often include microcephaly, cyclopia or synophthalmia, and a proboscis. Less-severely affected cases present with microcephaly, hypotelorism, midface hypoplasia, flat nasal bridges, cleft lip and/or palate, and/or a single maxillary incisor (Solomon et al. 2010b; Table 1.1, Figure 1.1). Additionally, midline defects such as undivided thalami, absent corpora callosa and absent or hypoplastic olfactory and optic bulbs and tracts occur (Solomon *et al.*, 2010b, 1993). A microform HPE also exists, where subtle facial phenotypes such as hypotelorism, a sharp and narrow nasal bridge and single maxillary incisor are present in the absence of structural brain abnormalities. These cases are often not recognized until a severely affected relative with HPE is identified (Solomon *et al.*, 2010b), and thus the frequency of microform HPE is underestimated.

1.1.2 MIHV HPE

In contrast to classic HPE, MIHV HPE (also known as syntelencephaly) presents with normal separation of the basal forebrain, anterior frontal lobes and occipital regions, but a failure to divide the posterior frontal and parietal regions of the cerebral hemispheres along the dorsal midline (Barkovich and Quint 1993; Simon et al. 2002; Lewis et al. 2002; Table 1.1, Figure 1.1). Additional structures, such as the caudate nuclei, thalami and mesencephalon can also be affected in MIHV cases. Whilst some similarities occur between classic and MIHV HPE (namely non-cleavage of a portion of the cerebral hemispheres), MIHV HPE is rarer and milder than classic HPE (Simon *et al.*, 2002), and likely has a distinct embryological origin (Fernandes et al. 2007).

1.2 The genetics of human HPE

When considered as a single disorder, HPE genetics exhibits extreme heterogeneity with multiple classes of causative mutations and numerous modes of heredity. Up to 50% of HPE cases are attributable to chromosomal abnormalities, while a further 25% of cases are syndromic and 25% occur in isolation. While some autosomal recessive cases of non-chromosomal, non-syndromic HPE have been reported, the condition is generally considered to be autosomal dominant (Barr and Cohen, 2002; Mercier *et al.*, 2011; Ming *et al.*, 2002; Mouden *et al.*, 2016; Roessler *et al.*, 2012b). HPE, however, does not exhibit simple Mendelian inheritance as evidenced by the facts that only ~70% of individuals who carry HPE pathogenic

mutations exhibit HPE symptoms (i.e. the condition is incompletely penetrant) (Mercier *et al.*, 2011) and that the same mutation can confer vastly different phenotypes in different carriers as a result of variable expressivity. There is evidence from human and animal studies that both genetic and environmental factors influence the HPE end-phenotype (Hong *et al.*, 2012; Kietzman *et al.*, 2014; Mouden *et al.*, 2016), therefore the ultimate consideration when studying HPE is the total activity achieved along the particular embryonic signalling pathways that direct hemisphere separation (Roessler and Muenke, 2010). Evidently, some cases of HPE have a digenic basis (Dubourg *et al.*, 2016; Mouden *et al.*, 2016). For the majority of HPE probands, however, the putative second hit factors remain unidentified and it seems a model of autosomal dominant with modifier effects most aptly describes HPE heritability (Odent *et al.*, 1998; Roessler *et al.*, 2012b).

One clear source of HPE genetic variability is locus heterogeneity (a single disorder is caused by mutations at different chromosomal loci). Fourteen genes have been implicated in non-syndromic classic and microform HPE in humans (*SHH, ZIC2, TGIF, SIX3, CDON, DISP1, DLL, FGF8, FGFR1, FOXH1, GAS1, PTCH1, NODAL, TDGF1)* (Arauz *et al.*, 2010; Bae *et al.*, 2011; Belloni *et al.*, 1996; Brown *et al.*, 2001, 1998; De la Cruz *et al.*, 2002; Dubourg *et al.*, 2016, 2007; Dupé *et al.*, 2011; Gripp *et al.*, 2000; Lacbawan *et al.*, 2009; Mercier *et al.*, 2011; Ming *et al.*, 2002; Pineda-Alvarez *et al.*, 2012; Ribeiro *et al.*, 2010; Roessler *et al.*, 2009a, 2009b, 2009c, 1996). These genes are classified as either 'large effect' (major) or 'small effect' (minor) HPE genes, according to how often they are mutated in the disorder. In classic HPE, the two genes most commonly mutated are *SHH* (12%) and *ZIC2* (9%), which together account for ~85% of solved probands (Dubourg *et al.*, 2016, 2011; Roessler *et al.*, 2009a). Additional genes have been associated with HPE in the mouse, indicating that they may be minor HPE genes and/or genetic modifiers in humans (reviewed in Schachter and Krauss 2008).

Notably, many genes associated with classic HPE do not cause MIHV HPE. In mice, *Zic2* and *Fgf8*, as well as the BMP ligands, BMP antagonists *Chordin* (*Chrd)* and *Noggin (Nog),* BMP receptors *Bmpr1a, Bmpr1b*, and the transcription factors *Lhx5* and *Rfx4,* have all been associated with MIHV (Anderson *et al.*, 2002; Cheng *et al.*, 2006; Dubourg *et al.*, 2016; Fernandes *et al.*, 2007; Lewis *et al.*, 2002; Nagai *et al.*, 2000; Simon *et al.*, 2002; Solomon *et al.*, 2010a; Storm *et al.*, 2006; Warr *et al.*, 2008). In contrast, relatively few genes (*ZIC2*, *FGF8*, deletion of *EYA4*) have been associated with MIHV HPE in humans (Abe *et al.*, 2009; Dubourg *et al.*, 2016; Solomon *et al.*, 2010a), of which *ZIC2* and *FGF8* are also associated with classic HPE. Animal models of *Zic2* dysfunction suggest that classic and MIHV HPE each have a distinct embryological basis, consistent with their observed differential involvement of the ventral versus dorsal brain regions

in human HPE. Thus, some phenotype variance in HPE occurs as a result of the disruption of related but distinct embryonic processes.

Despite recent progress in elucidating the genetic aetiology of human HPE, only 25% of cases of non-chromosomal, non-syndromic HPE have been attributed to pathogenic mutations in known HPE genes (Dubourg *et al.*, 2007; Roessler *et al.*, 2009a). Though clues from the patient's phenotype can, in some cases, point to a specific causative gene, generally, newly diagnosed HPE patients are screened for pathogenic deletions or point mutations via targeted gene panels that sequence the coding region only of the most common HPE-associated genes (SHH, ZIC2, SIX3 and TGIF) (Dubourg *et al.*, 2016; Solomon *et al.*, 1993). This restricted panel is due to logistical considerations, such as the limited availability of DNA (particularly in in cases where the patient is deceased) and the frequency with which mutations in these gene are identified in HPE cohorts. The remaining unsolved (non-chromosomal) cases are thought to be caused by environmental factors, unidentified HPE genes, and/or non-coding-region variants that alter the expression of known HPE-associated genes. An emerging area of HPE research aims to assess the likely contribution of putative risk factors (such as maternal diabetes, ethyl alcohol, cigarette smoking and retinoic acid) to the aetiology of HPE, either as single factors or in combination with pathogenic genetic lesions. Additionally, next-generation sequencing (NGS) targeted to 20 associated or candidate HPE genes and whole exome sequencing are currently being used to identify new causative HPE loci in probands and family members (Dubourg *et al.*, 2016; Mouden *et al.*, 2016). Whole exome sequencing, however, fails to account for variants in the non-coding regulatory regions of HPE-associated genes, which are a likely source for mutations in unsolved HPE cases. One such region, located 460 kb upstream of SHH, is known to control SHH expression in the developing forebrain via binding of the transcription factor SIX3. A rare variant in this region, identified in one proband with semilobar HPE, results in reduced SIX3/SIX6 binding affinity when analysed via EMSA in Cos-1 cells (Jeong *et al.*, 2008). This variant is hypothesised to result in reduced SHH expression in the ventral forebrain. Though this SHH regulatory region is included in current NGS panels, the identification of similar regulatory regions for other HPE-associated genes has remained elusive. **In this thesis, I will attempt to elucidate whether a similar region within the *ZIC2* 3'UTR regulates *Zic2* expression and contributes to HPE pathogenesis when mutated.**

Studies to date have already revealed that aspects of HPE pathogenesis vary according to the affected genetic locus, and thus, it is already possible to delineate several distinct features of *ZIC2*-associated HPE.

1.3 The *ZIC2* gene and protein

ZIC2, a member of the zinc finger of the cerebellum family of transcription factors (ZIC1-5), is arranged in a divergently transcribed bigene pair with *Zic5* on chromosome 13 in humans (14 in mice) and consists of three exons (Figure 1.2). Orthologues of the *Drosphila odd-paired (opa*), the ZIC proteins are defined by the presence of a zinc finger domain (ZFD) consisting of five tandem Cys2His2 zinc fingers (Aruga *et al.*, 2006)(Figure 1.2). The ZFD mediates DNA binding and protein-protein interactions in all family members, and is essential for the nuclear localisation of at least one ZIC protein (ZIC3) (Ware et al. 2004; Brown et al. 2005; Bedard et al. 2007; Hatayama et al. 2008; Pourebrahim et al. 2011, reviewed in Houtmeyers et al. 2013). A similar domain can be found in GLI, GLIS and NKL proteins. In addition to the ZFD, ZIC2 contains a zinc finger N-terminally conserved domain (ZF-NC), the ZIC/Odd paired conserved motif (ZOC) (Aruga *et al.*, 2006), as well as multiple polymeric amino acid tracts such as histidine, alanine and serine/glycine repeats. The presence of the ZOC domain and the makeup of the first zinc finger in ZIC2 results in its classification as a Subclass A ZIC protein, along with ZIC1 and ZIC3 (Houtmeyers *et al.*, 2013). The ZOC domain is required for protein-protein interactions, whilst expansion of the C terminal alanine tract results in reduced DNA binding activity despite the ZFD domain remaining functional, and has been linked to multiple HPE cases (Brown *et al.*, 2001; Himeda *et al.*, 2013; Mizugishi *et al.*, 2004). Recently, two additional domains were identified in ZIC1, 2 and 3: Subclass A N-terminally conserved domain (SANC) and Subclass A C-terminally conserved domain (SACC). Both domains are required for ZIC3 transactivation and presumably have a similar function in ZIC2 (Ahmed et al, manuscript under review).

During gastrulation *Zic2, Zic3* and *Zic5* are expressed in overlapping domains. It is this overlapping expression that is thought to contribute to functional redundancies between family members. Unique expression domains do exist, however, and it is these unique domains that have been attributed to the variety of diseases associated with Zic family members. At 5.5 dpc, *Zic2, Zic3* and *Zic5* are expressed in the ectoderm of the extra-embryonic and embryonic portions of the egg cylinder (Elms *et al.*, 2004; Furushima *et al.*, 2000). Transcripts have also been identified in the mesoderm that ingresses through the primitive streak, however expression is absent from the mesoderm that migrates to the extraembryonic regions. The distinctive expression of *Zic2* begins at 7.0 dpc when it can be detected in the embryonic node (Figure 1.3). Though *Zic3* is also expressed in the node, this does not occur until late gastrulation-early somite stages (7.75 - 8.0 dpc) after *Zic2* node expression has ended (Elms *et al.*, 2004). As the mesoderm ingresses into the primitive streak, *Zic2* transcripts recede from the posterior primitive streak until expression in the primitive streak is ceased at the early headfold stage of gastrulation (7.75 dpc). It is during the establishment of the neural plate that *Zic2* recedes from

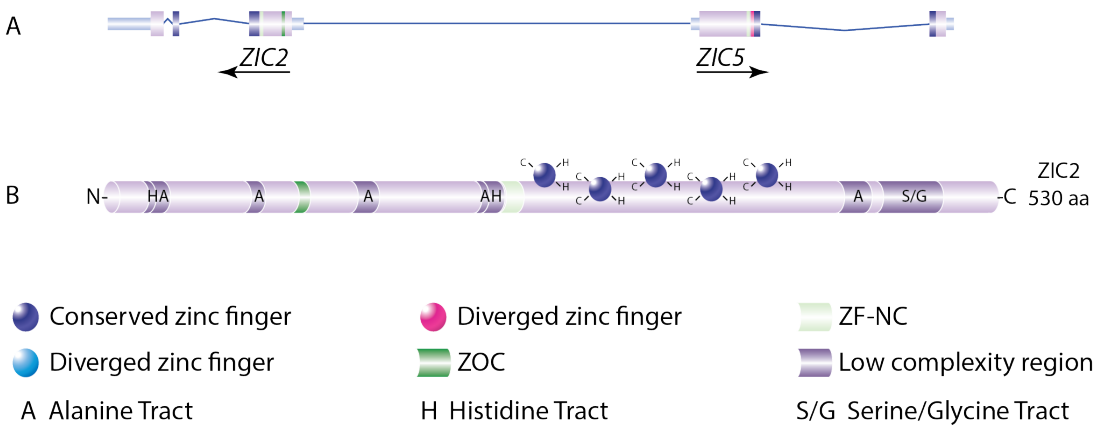


Figure 1.2: The location and structure of ZIC2. (a) *ZIC2* is arranged in a divergently transcribed pair with *ZIC5* on chromosome 13 in humans (14 in mice), with three exons and two introns. **(b)** Like the other ZIC family members, the ZIC2 protein contains a highly conserved ZF-NC domain followed by five tandem C_2H_2 zinc finger domains which provide DNA binding activity. A ZOC domain can also be found in ZIC2 towards the N-terminal, along with several low complexity regions such as alanine and histidine tracts and serine/glycine tracts. Figure is modified from (Houtmeyers et al. 2013).

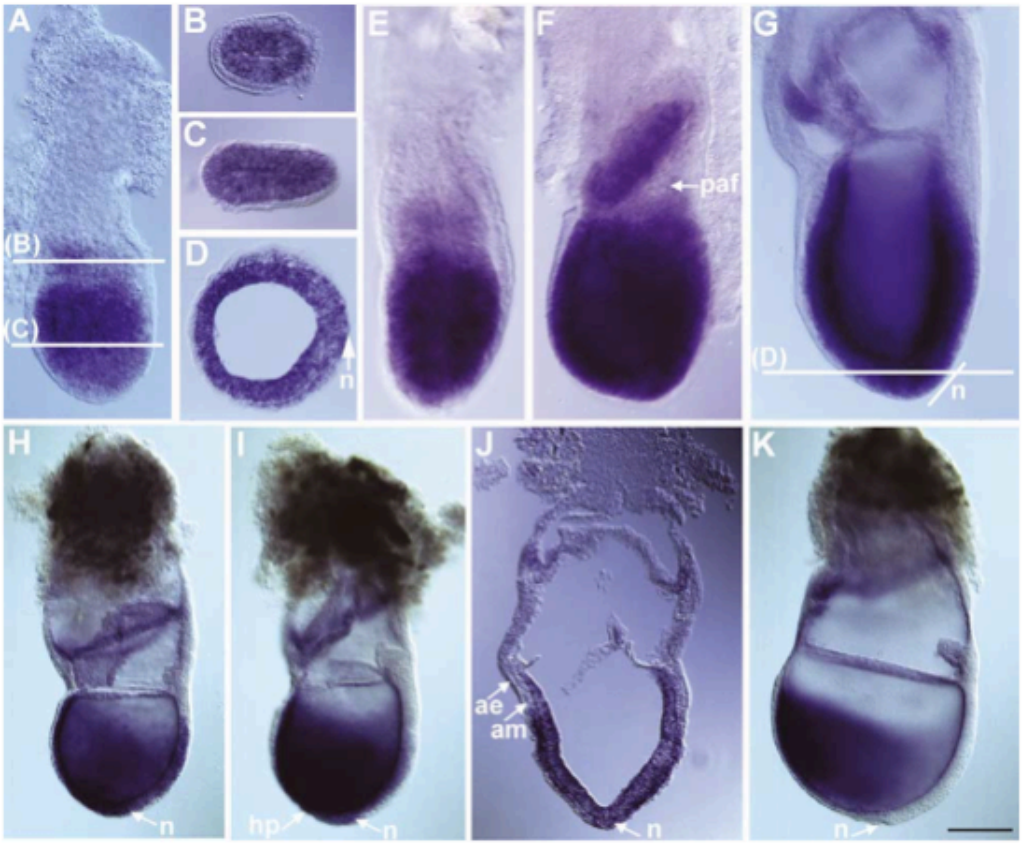


Figure 1.3: *Zic2* expression during mouse gastrulation. (a) A 5.5 dpc, pre-streak stage embryo. **(b)** A transverse section through the extra embryonic region of a pre-streak stage embryo at the level shown in (a). **(c)** A transverse section through the embryonic region of a pre-streak stage embryo at the level shown in (a). **(d)** A transverse section through the embryonic region of a 7.0 dpc, late-streak stage embryo at the level shown in (g). The primitive streak and node are to the right. **(e)** A 6.5 dpc, early-streak stage embryo at the onset of gastrulation. **(f)** A 6.75 dpc mid-streak stage embryo with mesoderm moving into the posterior amniotic fold. **(g)** A 7.0 dpc late-streak stage embryo. The node is now visible and *Zic2* transcripts are seen in the node. **(h)** A 7.25 dpc early allantoic bud stage embryo. **(i)** A 7.5 dpc late allantoic bud stage embryo. Expression in the node and emerging head process is still seen. **(j)** A longitudinal section through a 7.5 dpc embryo such as that shown in (i). *Zic2* transcripts are not found in the mesoderm of the extra embryonic region or of the proximal embryonic region. **(k)** A 7.75 dpc early head-fold stage embryo. *Zic2* transcripts are now mainly confined to the anterior half of the embryo and the expression in the node has ceased. ae: anterior definitive endoderm, am: anterior mesoderm, hp: head process, n: node, paf: posterior amniotic fold. Scale bar, 50 mm (a – d), 100 mm (e – g), 200 mm (h, i and k) and 170mm (j). Figure taken from (Elms *et al.*, 2004) with permission.

the posterior embryo proper and becomes restricted to the anterior neuroectoderm. By this time, *Zic3* and *Zic5* are also restricted to this region. The expression of the *Zics* in the neuroectoderm becomes limited to the dorsal region, which will eventuate into neural crest cells and dorsal neuron production (Elms *et al.*, 2004; Houtmeyers *et al.*, 2013). After the neural tube closes, high levels of *Zic2* expression can be detected in the dorsal telencephalon (roof-plate and hippocampal primordium) (Cheng *et al.*, 2006; Okada *et al.*, 2008) and in the anterior and posterior ventral telencephalon (Okada *et al.*, 2008). It is postulated that loss of *Zic2* expression in the mid-gastrula node and developing forebrain results HPE.

1.4 The genetics of *ZIC2*-associated HPE

One unusual aspect of ZIC2 involvement in HPE is that probands with *ZIC2* mutations have been found across the entire HPE phenotypic spectrum. Generally, classic HPE genes are not associated with MIHV. Although the vast majority of *ZIC2* mutations result in classic HPE, a few are associated with MIHV (Table 1.2). A striking demonstration of the variable expressivity of *ZIC2* mutations comes from the report of monozygotic twins with the same *de novo* mutation at a splice donor site in *ZIC2* (c.1239+1G>C). These twins exhibit different classes of HPE: one twin developed semilobar HPE whilst the other had MIHV (Nakayama *et al.*, 2016). As discussed later, animal models suggest the classic and MIHV forms of HPE have a distinct embryological basis and it appears that ZIC2 is required for both processes in man and mice. In comparison to HPE as a whole, *ZIC2*-associated HPE manifests as highly penetrant (93%) with relatively few mild phenotype individuals (Solomon *et al.*, 2010a). In fact, it is estimated that 90% of patients with a *ZIC2* mutation exhibit structural brain anomalies (Solomon *et al.*, 2010b). Additionally, *ZIC2* mutation is more frequently associated with sevre structural brain anomalies, (alobar or semilobar), which account for 75% of the *ZIC2*–associated HPE cases in which phenotype class is recorded (Table 1.2). In contrast, a similar analysis of 92 individuals (probands and family members) with clinically apparent HPE and *SHH* mutations found that 39% exhibited Alobar or Semilobar brain abnormalities, that 48% presented with no brain abnormalities but had craniofacial characteristics of microform HPE and none had MIHV (Solomon *et al.*, 2010b). *ZIC2* mutation therefore accounts for the majority of severely affected HPE cases (Solomon *et al.*, 2010b).

Another striking observation based on the molecular subtyping of HPE cases is that *ZIC2* mutation breaks the mantra 'the face predicts the brain', with the craniofacial defects typically associated with classic HPE absent in those patients assessed (Brown *et al.*, 1998; Solomon *et al.*, 2010a, 2010b). In particular, Solomon et al (2010a) found no *ZIC2*-associated HPE case presenting with facial findings at the severe end of the spectrum (cyclopia,

Table 1.2: Frequency of known HPE mutations in human *ZIC2* arranged by mutation type and phenotype. Polymorphisms were not included. Identical mutations that occurred in multiple patients were counted as independent instances. Mutation data was collated from (Dubourg *et al.*, 2016; Mercier *et al.*, 2011; Nakayama *et al.*, 2016; Paulussen *et al.*, 2010; Ribeiro *et al.*, 2012; Roessler *et al.*, 2012a, 2012b, 2009a; Solomon *et al.*, 2010a). Mic: microform; MIHV: middle interhemispheric variant; SNV: single nucleotide variant.

	Unknown	**Alobar**	**Semilobar**	**Lobar**	**Mic**	**MIHV**
Missense	18 (10.34%)	7 (4.02%)	7 (4.02%)	4 (2.30%)	2 (1.15%)	0 (0.00%)
Nonsense	7 (4.02%)	5 (2.87%)	8 (4.60%)	2 (1.15%)	1 (0.57%)	0 (0.00%)
Frameshift	19 (10.92%)	10 (5.75%)	23 (13.22%)	2 (1.15%)	5 (2.87%)	0 (0.00%)
Insertion	1 (0.57%)	0 (0.00%)	1 (0.57%)	0 (0.00%)	0 (0.00%)	0 (0.00%)
Deletion	2 (1.15%)	1 (0.57%)	1 (0.57%)	1 (0.57%)	0 (0.00%)	1 (0.57%)
Duplication	8 (4.60%)	5 (2.87%)	9 (5.17%)	1 (0.57%)	1 (0.57%)	3 (1.72%)
Splice variant (intron)	6 (3.45%)	2 (1.15%)	3 (1.72%)	0 (0.00%)	0 (0.00%)	1 (0.57%)
SNV (3'UTR)	3 (1.72%)	1 (0.57%)	1 (0.57%)	0 (0.00%)	2 (1.15%)	0 (0.00%)
Total (n=174)	**64 (36.78%)**	**31 (17.82%)**	**53 (30.46%)**	**10 (5.75%)**	**11 (6.32%)**	**5 (2.87%)**
Total (n=110)		**31 (28.18%)**	**53 (48.18%)**	**10 (9.09%)**	**11 (10.00%)**	**5 (4.55%)**

synophthalmia, or a proboscis) or a combination of facial features similar to those caused by mutations in other HPE genes. Instead, a distinct phenotype can be seen in *ZIC2*-associated HPE, consisting of (but not always containing) bitemporal narrowing, upslanted palprebral fissures, a flat nasal bridge, short nose with anteverted nares, a broad and deep philtrum and the appearance of large ears. Whilst facial clefts occurred in both non-*ZIC2* and *ZIC2*-associated HPE, the frequency is reduced by 1/3 in the latter (Solomon *et al.*, 2010a). Despite the mild facial phenotype, these patients often have severe HPE and neurologic impairment (Solomon *et al.*, 2010a, 2010b).

Additionally, other non-forebrain phenotypes are often associated with *ZIC2*-HPE. For example, HPE patients with intragenic *ZIC2* mutations have been found to exhibit neural tube defects (4%), hydrocephalus (12%), skeletal anomalies (14%), cardiac anomalies (9%) and renal anomalies (7%) (Solomon *et al.*, 2010a). Mouse models that recapitulate *ZIC2*-associated HPE also display a subset of these co-morbidities. Both severe and mild *Zic2* loss-of-function leads to incompletely penetrant neural tube defects such as exencephaly and spina bifida, due to a requirement for ZIC2 during neurulation (Elms *et al.*, 2003; Nagai *et al.*, 2000; Ybot-Gonzalez *et al.*, 2007). In the severe loss-of-function allele of murine *Zic2, Kumba (Ku)*, spina bifida occurs due to the downregulation of BMP antagonists such as Noggin in the dorsal neural plate. These antagonists are required to induce the formation of dorsolateral hinge points (DLHPs) in the neural tube of the lower spinal neuraxis. In $Zic2^{Ku/Ku}$ embryos, DLHPs are absent from this region (Ybot-Gonzalez *et al.*, 2007). The same mechanism is presumed to be at play in hypomorphic *Zic2* mouse mutants ($Zic2^{tm1Jaru}$ MGI:2156825; aka *kd*), where abnormal folding of the posterior neural tube at 9.5 dpc leads to spina bifida (Nagai *et al.*, 2000). Consequently, these *Zic2* mutants provide an opportunity to model additional co-morbidities of *Zic2*-associated HPE. **In this thesis, I will investigate whether $Zic2^{Ku/Ku}$ embryos also exhibit cardiac anomalies.**

Comparison of relatively large HPE cohorts has also revealed aspects of *ZIC2*-HPE heritability. For example, family analysis has shown that *ZIC2* mutations are largely *de novo*, with an inheritance rate of only 27-30% (Mercier *et al.*, 2011; Mouden *et al.*, 2016; Solomon *et al.*, 2010b). In contrast, 70% of *SHH* and *SIX3* mutations are inherited (Solomon *et al.*, 2010a). Of the inherited *ZIC2* cases, two-thirds were maternally inherited and one-third paternally inherited. Additionally, families with *ZIC2* mutations in greater than two generations have not been reported. A small subset of *ZIC2*-associated HPE cases, however, appear to occur due to allelic drop out or germline mosaicism, resulting in parents negative for mutations siring multiple affected children (Solomon *et al.*, 2010a, 2010b). The low rate of inheritance in *ZIC2* cohorts may be due to severely affected individuals being unable to reproduce (Solomon *et al.*, 1993), and suggests that mutations in *ZIC2* produce a higher level of lethality in comparison to other HPE

genes. Despite this, a recent analysis by Weiss et. al., found that individuals with *ZIC2*-associated HPE were more likely to survive to adolescence than those with HPE due to mutations in other implicated genes (Weiss *et al.*, 2017). It remains unclear from the targeted sequencing analysis to date whether heterozygous *ZIC2* mutation is sufficient to cause HPE. There are two reported cases of *ZIC2* CDS mutations in conjunction with other major HPE genes (*ZIC2/SHH and ZIC2/SIX3*), and mutations in the 3'UTR of *ZIC2* have also been found in conjunction with mutations in other HPE genes (Lacbawan *et al.*, 2009; Nanni *et al.*, 1999; Roessler *et al.*, 2012a). Studies of other HPE-associated genes show that, in mice, exposure to alcohol during gestation in conjunction with a pre-existing mutation increases the frequency and severity of HPE cases (Aoto *et al.*, 2008; Hong *et al.*, 2012; Kietzman *et al.*, 2014). It is therefore possible that at least some cases of HPE arise when a *ZIC2* mutation sensitises the developing embryo to one or multiple teratogens. The involvement of gene x gene and gene x environment interactions may explain why for *ZIC2* (and other HPE associated genes), heterozygous mutations give rise to the HPE phenotype in man but not mouse (Brown *et al.*, 2005; Chiang *et al.*, 1996; Nagai *et al.*, 2000; Petryk *et al.*, 2015; Roessler *et al.*, 2009a; Schachter and Krauss, 2008; Warr *et al.*, 2008).

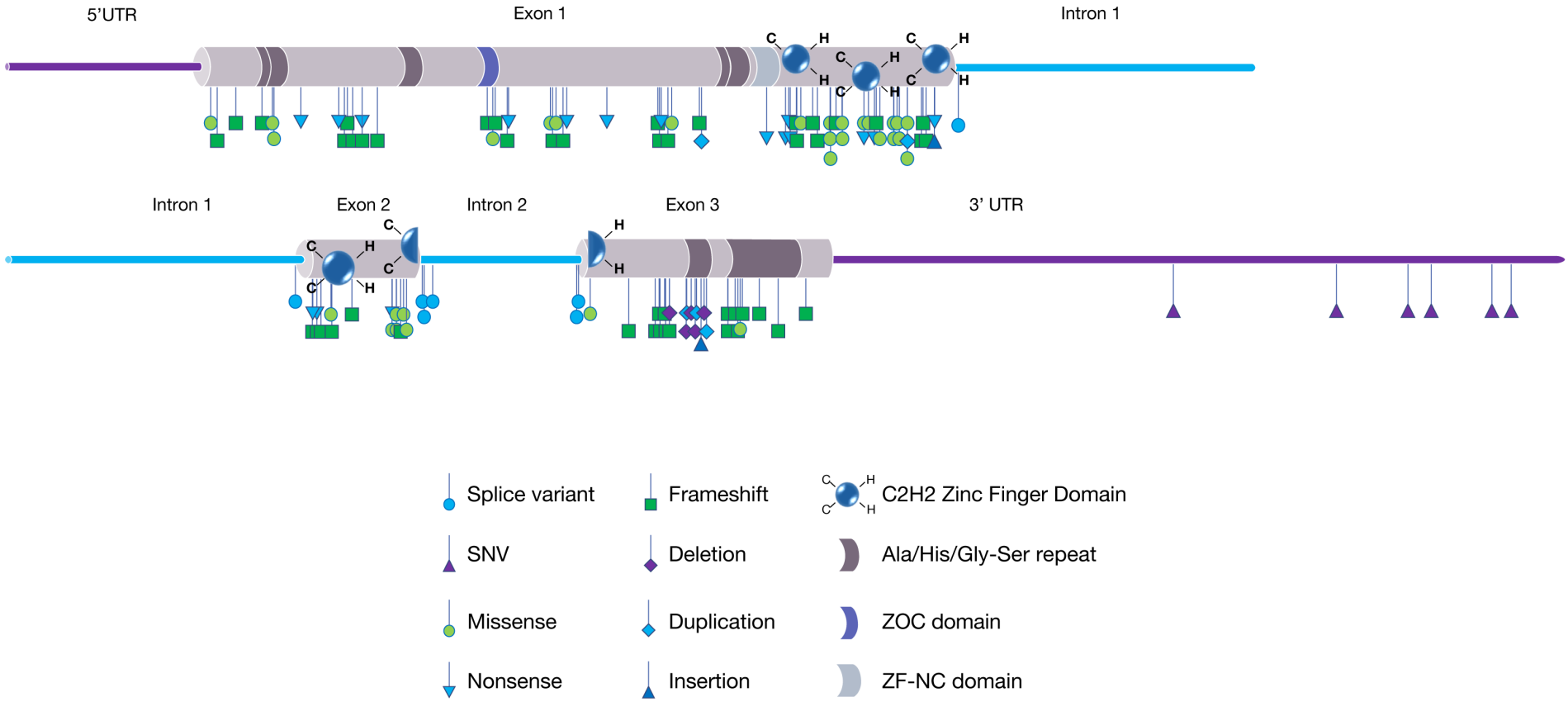
This analysis of HPE cohorts has enabled assessment of the *ZIC2* mutational spectrum, confirming that *ZIC2*-associated HPE exhibits allelic heterogeneity (whereby different mutations in the same gene give rise to the same disease). In the last 12 years the number of published unique *ZIC2*-associated HPE mutations has grown from 20 (Brown *et al.*, 2005) to 118 at the time of this thesis (Table 1.3), and will continue to expand as sequencing techniques become more readily available and the unique *ZIC2*–associated HPE phenotype is refined. Among the 105 documented HPE cases with mutations in the *ZIC2* coding sequence, the majority are predicted to substantially alter the *ZIC2* transcript (44.76% are frameshift, 14.29% nonsense, 4.76% duplication, 4.76% deletion and 2.86% insertion) rather than a single amino acid (28.57% missense) (Table 1.3, Figure 1.4). Similarly, of the 13 mutations in *ZIC2* non-coding DNA, 53.85% are splice variants and predicted to substantially alter the transcript. Consistent with the mutational landscape of *ZIC2*, analysis by Solomon *et al* (2010a) found that 98% of all *ZIC2* mutations were predicted or proven to be loss-of- function.

Analysis of mutation type and location underscores the functional importance of the ZFD. For example, 45.71% of all *ZIC2* mutations occur in the ZFD (Table 3) and 73% (22/30) of ZFD mutations are missense, demonstrating that single amino acid changes in the ZFD are sufficient to cause disease. Furthermore, no nonsense mutations, nor most frameshift mutations, produce

Table 1.3: Type and frequency of known HPE mutations in human *ZIC2*. Polymorphisms were not included. Identical mutations that occurred in multiple patients were counted as one instance. Mutation data was collated from (Dubourg *et al.*, 2016; Mercier *et al.*, 2011; Nakayama *et al.*, 2016; Paulussen *et al.*, 2010; Ribeiro *et al.*, 2012; Roessler *et al.*, 2012a, 2012b, 2009a; Solomon *et al.*, 2010a). SNV: single nucleotide variant, CDS: coding DNA sequence.

	Number	**Percent of CDS mutations**	**Percent of non-coding mutations**	**Percent of all mutations**
CDS				
Missense	30	28.57%		25.42%
Nonsense	15	14.29%		12.71%
Frameshift	47	44.76%		39.83%
Duplication	5	4.76%		4.24%
Insertion	3	2.86%		2.54%
Deletion	5	4.76%		4.24%
Total	105			
Non-coding DNA				
Splice variant (intron)	7		53.85%	5.93%
SNV (3'UTR)	6		46.15%	5.08%
Total	13			118

Figure 1.4: The genomic and protein structure of ZIC2, showing the known human HPE-associated variants. Polymorphisms were not included. Identical mutations that occurred in multiple patients were counted as one instance. Mutation data was collated from (Dubourg *et al.*, 2016; Mercier *et al.*, 2011; Nakayama *et al.*, 2016; Paulussen *et al.*, 2010; Ribeiro *et al.*, 2012; Roessler *et al.*, 2012a, 2012b, 2009a; Solomon *et al.*, 2010a). SNV: single nucleotide variant, ZOC: ZIC/Odd paired conserved motif, ZF-NC: Zinc finger N-terminally conserved domain.


5'UTR
Exon 1
Intron 1
Intron 1
Exon 2
Intron 2
Exon 3
3' UTR
Splice variant
SNV
Missense
Nonsense
Frameshift
Deletion
Duplication
Insertion
C2H2 Zinc Finger Domain
Ala/His/Gly-Ser repeat
ZOC domain
ZF-NC domain

a complete ZFD when translated (Roessler *et al.*, 2009a). In contrast, few missense variants occur outside of the ZFD where 86% of known variants are predicted to drastically alter the *ZIC2* transcript. Another notable enrichment in mutation type is found in the C-terminal alanine repeat (Table 1.4), a region that influences the strength of ZIC2 DNA binding and transcriptional activity. Eight duplication, deletions or insertions have been identified in this small (15 amino acid) region out of thirteen variants (62%) of this type across the whole gene. Functional analysis of these variants shows that expansion from 15As to 25As results in near-complete loss of transactivation, yet reduction to 2As results in both an increase and decrease in transactivation, dependant on the promoter that is used. This indicates that the alanine tract can modulate ZIC2 transactivation, contingent on the DNA sequence being targeted (Brown *et al.*, 2005). Expansion of the *ZIC2* alanine repeat to 25A has been identified in multiple unrelated families, and in some cases is hypothesized to occur via errors in somatic recombination in the patient's fathers (Brown *et al.*, 2001).

The ongoing efforts to delineate genotype-phenotype correlations in HPE have shed light on the genetic and embryonic origin of HPE and, with respect to aetiology, have led to the following generalisations:

- HPE arises as a consequence of failed dorsal-ventral (D-V) forebrain patterning.
- Classic HPE arises due to failed ventral patterning, whereas MIHV HPE is associated with failed dorsal patterning.
- Classic HPE brain abnormalities can occur with or without associated facial abnormalities.
- For some molecular subclasses of HPE, classic or MIHV HPE brain abnormalities can occur alongside abnormalities not associated with the face and forebrain (e.g. *Zic2*).

To consider how *ZIC2* mutation causes the typical brain abnormalities of HPE requires understanding how D-V forebrain pattern is established. Much of our knowledge regarding mammalian embryonic development and the mechanism of ZIC2 activity during brain development is derived from the analysis of mouse mutant phenotypes and the next sections will refer to mouse development.

1.5 Ventral forebrain patterning during murine development

The generation of a correctly patterned embryo is dependent upon inductive interactions between progenitor tissues that direct differentiation, regionalisation and morphogenesis. In the ventral forebrain, patterning is a consequence of inductive interactions between the

Table 1.4: Protein location and frequency of known HPE mutations in the human ZIC2 protein. Polymorphisms and chromosome deletions were not included. Identical mutations that occurred in multiple patients were counted as one instance. Mutation data was collated from (Dubourg *et al.*, 2016; Mercier *et al.*, 2011; Nakayama *et al.*, 2016; Paulussen *et al.*, 2010; Ribeiro *et al.*, 2012; Roessler *et al.*, 2012b, 2009a; Solomon *et al.*, 2010a). ZOC: ZIC/Odd paired conserved motif, ZF-NC: Zinc finger N-terminally conserved domain.

Protein Domain	Location (aa)	Missense	Nonsense	Frameshift	Duplication	Insertion	Deletion	Total (n=105)
Histidine Repeat	20-24							**0 (0.00%)**
Alanine Repeat	24-34			1				**1 (0.95%)**
Alanine Repeat	89-98							**0 (0.00%)**
ZOC	115-126			1				**1 (0.95%)**
Alanine Repeat	226-231							**0 (0.00%)**
Histidine Repeat	231-240							**0 (0.00%)**
ZF-NC	241-256		1					**1 (0.95%)**
C2H2 Zinc Finger Domain	258-416	22	8	16	1	1		**48 (45.71%)**
Alanine Repeat	456-471				3	1	4	**8 (7.62%)**
Serine/Glycine Repeat	477-516	1		8				**9 (8.57%)**
Non-domain		7	6	21	1	1	1	**37 (35.24%)**
Total (n=105)		**30 (28.57%)**	**15 (14.29%)**	**47 (44.76%)**	**5 (28.57%)**	**3 (2.86%)**	**5 (4.76%)**	

anterior neurectoderm and the underlying prechordal plate (PrCP) (reviewed in Placzek and Briscoe 2005). Both of these tissues arise during gastrulation (reviewed in Arkell and Tam 2012), prior to which the neurectoderm precursor cells are located within the distal epiblast where they are encased by the distal visceral endoderm cells (Figure 1.5A). These visceral endoderm (VE) cells secrete factors that antagonize TGF-β and WNT signalling activities, protecting the distal ectoderm cells of the epiblast from the signals that would otherwise cause them to differentiate. By the time gastrulation begins, the future neurectoderm cells are found at the anterior of the epiblast where they continue to be protected from differentiation into endoderm or mesoderm by antagonists secreted from the enveloping endoderm, which at this point is called the anterior visceral endoderm (AVE). During gastrulation, the anterior neurectoderm population expands anteriorly and proximally and occupies two-thirds of the ectoderm layer by mid-gastrulation (Figure 1.5B, 7.0 days post-coitum; dpc). Although the descendants of some of these cells will colonise more posterior parts of the brain, the cells in this region are becoming progressively restricted to forebrain fate. Approximately half of the neurectoderm that has been produced at this stage will give rise to the forebrain, therefore perturbations in neural development at this stage disproportionately affect the forebrain.

The tissues that will form the PrCP are initially co-localised with the precursors of the definitive endoderm, and prior to gastrulation are found at the future posterior side of the embryo about halfway along the embryonic portion of the epiblast (Figure 1.5A). During gastrulation, these cells ingress through the anterior segment of the primitive streak and extend along the embryonic midline to reach the entire length of the body axis. The first cells to pass through the primitive streak will come to lie at the embryonic anterior, with the later cells taking up an axial position in accordance with their time of passage through the primitive streak. When the cells first emerge from the primitive streak and migrate along the midline, they do so as a contiguous sheet of cells in which the midline mesoderm is flanked by definitive endoderm. At this stage, this sheet of cells is named the anterior mesendoderm (AME). Later in development, the cells at the midline separate and take up a position among the mesoderm tissues. The resulting axial mesoderm structures are named for their position along the body axis: the axial mesoderm that underlies the forebrain is called the PrCP and that which associates with the rest of the brain is the anterior notochord.

Clues to the molecular nature of the ventral forebrain patterning signals are provided by the resulting phenotypes in murine embryos with particular genetic mutations. Classic HPE arises following the loss of morphogenetic signalling activity induced by disruption of the HH ligand SHH, or combined mutation of the TGF-β antagonists *Chrd* and *Nog* in murine embryos

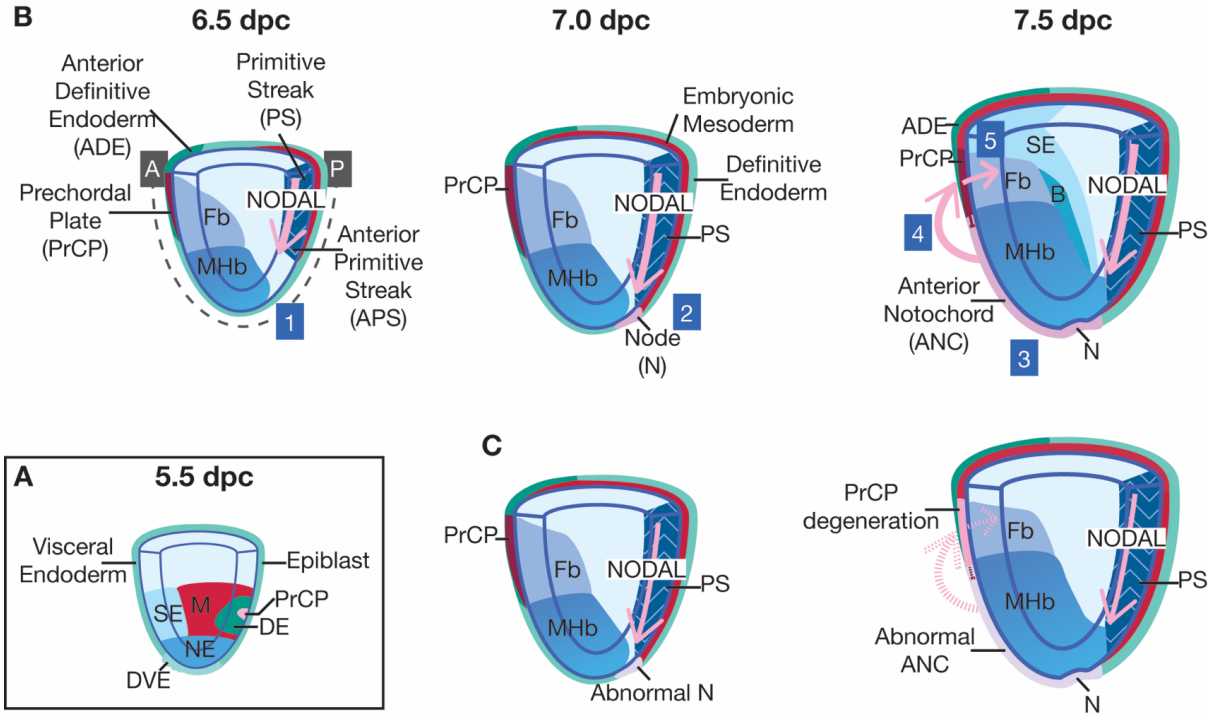

B
6.5 dpc
7.0 dpc
7.5 dpc
Anterior Definitive Endoderm (ADE)
Primitive Streak (PS)
A
P
Prechordal Plate (PrCP)
Fb
NODAL
MHb
Anterior Primitive Streak (APS)
1
Embryonic Mesoderm
Definitive Endoderm
PrCP
PS
2
Node (N)
ADE
PrCP
5
SE
Fb
B
NODAL
4
MHb
PS
Anterior Notochord (ANC)
3
N
A
5.5 dpc
Visceral Endoderm
Epiblast
SE
M
PrCP
DE
NE
DVE
C
PrCP degeneration
Abnormal ANC
Abnormal N

Figure 1.5: The origin of the prechordal plate and other tissues involved in dorsal-ventral patterning of the murine telencephalon. (a) A cut-away diagram of the pre-gastrula mouse embryo (5.5 dpc) with a superimposed cellular fate map. The pre-gastrula embryo is bi-laminar, with the inner ectoderm tissue (the epiblast) enveloped by the visceral endoderm. The position of the precursor cells for the ectoderm (neural; NE and surface; SE), mesoderm (M), definitive endoderm (DE) and anterior prechordal plate (PrCP) is shown. **(b)** Cut-away diagrams of the embryonic portion of the early (6.5 dpc), mid (7.0 dpc) and late (7.5 dpc) gastrulas. The anterior (A) and posterior (P) of the embryo are marked and the dotted line indicates the A-P axis. Wild type embryos use five steps to establish the ventral signalling centre in the forebrain neurectoderm. Step 1: NODAL signal in the posterior directs the initial differentiation of the PrCP and anterior definitive endoderm cells which transit the anterior primitive streak and migrate to the anterior midline of the embryo. Step 2: NODAL signal induces the transition of the anterior primitive streak cells into the overt node. Step 3: the anterior notochord cells transit the node and migrate to the anterior midline, coming to lie caudal of the PrCP. Step 4: inductive interactions between the PrCP and anterior notochord (negative from PrCP to ANC and positive from ANC to PrCP) stabilise PrCP fate enabling SHH secretion. Step 5, SHH signals vertically to the overlying forebrain neurectoderm to establish *Shh* expression in the rostral ventral neural midline and overlay ventral identity information on the neurectoderm which initially is dorsal in character. **(c)** The alterations in PrCP development in *Zic2*$^{Ku/Ku}$ embryos are shown. At 6.5 dpc *Zic2*$^{Ku/Ku}$ embryos are indistinguishable from wild type and Step 1 proceeds as normal. At 7.0 dpc, the lack of functional ZIC2 alters the level of perceived NODAL signal such that the overt node is formed (Step 2), but gene expression at the node is abnormal. Step 3 fails in the *Zic2*$^{Ku/Ku}$ mutants and by 7.5 dpc there is no anterior notochord. Consequently, the PrCP degenerates, does not secrete SHH and *Shh* expression does not initiate in the rostral ventral neural midline, resulting in classic HPE. dpc: days post-coitum, B: neural plate border, DE: definitive endoderm, DVE: distal visceral endoderm, Fb: forebrain neurectoderm, M: mesoderm, MHb: mid- and hindbrain neurectoderm, NE: neurectoderm, SE: surface ectoderm.

(Anderson *et al.*, 2002; Chiang *et al.*, 1996). This suggests that the HH pathway must be activated and the TGF-β pathway inhibited for ventral forebrain patterning to proceed. In the HH pathway (Figure 1.6B), the SHH ligand, along with co-receptors CDON, BOC, LRP2 and GAS1, binds to Patched (PTCH1), a transmembrane receptor. This interaction with PTCH1 relieves inhibition of Smoothened (SMO), which then facilitates the production and nuclear transport of full length, activating forms of GLI transcription factors to promote transcription of SHH targets (reviewed in Xavier et al. 2016). As described previously, mutation of not only *SHH* itself but also components of the HH transduction pathway have been found to be mutated in HPE probands, indicating HH signalling is also required for human forebrain ventral patterning. *Shh* transcripts are initially found in the PrCP itself and subsequently in the overlying rostral-ventral neural midline (RVNM). The downstream components and target genes of the SHH transduction pathway are expressed in the RVNM. Mutations in the response components of the SHH pathway produce neurectoderm that is incompetent to respond to the SHH signal, resulting in HPE (Fuccillo *et al.*, 2004; Spoelgen *et al.*, 2005).

In the TGF-β pathway (Figure 1.6A and C), ligands bind to and activate a Type I and Type II receptor complex, causing phosphorylation of some members of the SMAD family called the receptor-associated SMADs (R-SMADS). This enables their interaction with the common mediator SMAD (Co-SMAD) resulting in nuclear localization, the formation of higher order transcriptional complexes and regulation of TGF-β target genes. In contrast, BMP ligands signal through different receptors and SMAD molecules than the NODAL and GDF molecules, discussed below. Both gain- and loss-of-function experiments in the mouse indicate that BMP signalling represses *Shh* expression in the RVNM (Anderson *et al.*, 2002). The BMP antagonists *Chrd* and *Nog* are expressed in the node and axial mesoderm derivatives (notochord and PrCP) and in $Chrd^{-/-};Nog^{+/-}$ mutant embryos, BMP antagonism is reduced, resulting in embryos that exhibit cyclopia and HPE in conjunction with loss of *Shh* expression in the PrCP (Anderson *et al.*, 2002). Counterintuitively, *Bmp7* is expressed in the node, notochord and caudal PrCP alongside *Chrd* and *Nog* (Anderson *et al.*, 2002; Arkell and Beddington, 1997). Here, BMP7 may (as in the chick) modify the response of ventral midline cells to SHH and induce a rostral identity, instead of a floor-plate identity that would be induced by SHH on its own (Dale *et al.*, 1997). Unlike *Bmp7*, *Bmp2* and *Bmp4* are expressed in the surface ectoderm and, in the case of *Bmp2*, paraxial mesoderm adjacent to the PrCP, but not in the PrCP itself (Anderson *et al.*, 2002). Mutations in the BMP part of the TFG-β pathway have not yet been associated with human HPE, perhaps because mutations that result in elevated signalling are relatively rare compared to those that cause loss of signalling.

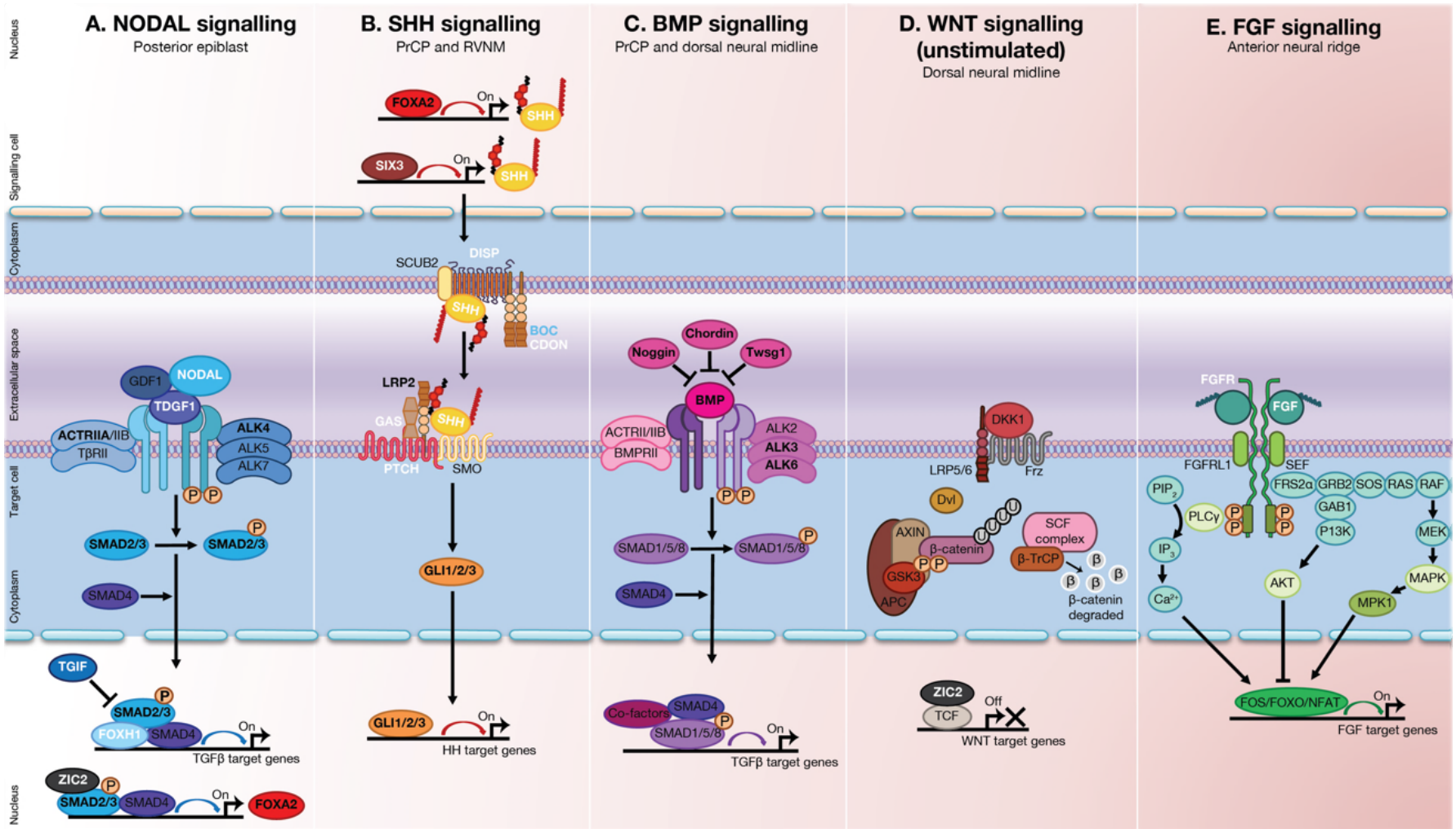

A. NODAL signalling
Posterior epiblast
B. SHH signalling
PrCP and RVNM
C. BMP signalling
PrCP and dorsal neural midline
D. WNT signalling (unstimulated)
Dorsal neural midline
E. FGF signalling
Anterior neural ridge
Nucleus
Signalling cell
Cytoplasm
Extracellular space
Target cell
Cytoplasm
Nucleus
FOXA2
On
SHH
SIX3
SCUB2
DISP
BOC
CDON
LRP2
GAS
PTCH
SMO
GLI1/2/3
Noggin
Chordin
Twsg1
BMP
ACTRII/IIB
BMPRII
ALK2
ALK3
ALK6
SMAD1/5/8
SMAD4
Co-factors
GDF1
NODAL
TDGF1
ACTRIIA/IIB
TβRII
ALK4
ALK5
ALK7
SMAD2/3
TGIF
FOXH1
TGFβ target genes
HH target genes
ZIC2
FOXA2
DKK1
LRP5/6
Frz
Dvl
AXIN
β-catenin
GSK3
APC
SCF complex
β-TrCP
β-catenin degraded
Off
TCF
WNT target genes
FGFR
FGF
FGFRL1
SEF
FRS2α
GRB2
SOS
RAS
RAF
GAB1
P13K
PIP2
PLCγ
IP3
Ca2+
AKT
MEK
MAPK
MPK1
FOS/FOXO/NFAT
FGF target genes

Figure 1.6: Signalling pathways involved in dorsal-ventral patterning of the murine telencephalon. (a) Nodal signalling. Mature Nodal ligands complex with the EGF-CFC co-factor TDGF1 (Cripto), Type I receptors (ALK4/5/7) and Type II receptors (ActRII or ActRIIB). Receptor activation leads to the phosphorylation of the type I receptor by the type II kinase, as well as phosphorylation of SMAD2 or SMAD3, which dimerize with SMAD4. The SMAD2/3-4 complex translocates to the nucleus and interact with the transcription factor FOXH1 and promote transcription of TGFβ target genes, or with ZIC2 to promote transcription of FOXA2. **(b) SHH signalling**. In the absence of ligand, the transmembrane domain protein Patched1 (PTCH) inhibits the activity of Smoothened (SMO) in the target cell. In a signalling cell, SHH expression is initiated by the transcription factors FOXA2 or SIX3 binding to enhancers. In the extracellular space, the binding of secreted SHH to PTCH releases SMO, allowing for regulation of HH target genes by the transcription factors GLI1, GLI2 and GLI3. **(c) BMP signalling.** The secreted BMP ligand binds to Type I (ALK2/3/6) and Type II receptors (ACTRII and ACTIIB). This results in the phosphorylation of DNA binding proteins SMAD1/5/8, which complex with SMAD4. The SMAD1/5/8-4 complex translocates to the nucleus and via interactions with co-factors, regulates transcription of TGFβ target genes they regulate transcription of BMP target genes. In the extracellular space, the secreted antagonists Chordin (CHRD) and Noggin (NOG) directly interact with BMP ligands to prevent the activation of downstream effectors whilst twisted gastrulation (TWSG1) can act as both an antagonist or agonist of BMP signalling in a cells-specific manner. **(d) WNT signalling.** In the absence of ligand/presence of inhibitors such as DKK1, β-catenin is phosphorylated (P) by the kinase activity of the destruction complex (consisting of Axin, APC, and GSK3), polyubiquitylated (U) by the SCF (SKP1, Cullin, F-box)/BTrCP complex and degraded by the proteasome. ZIC2 acts as a transcriptional co-repressor, complexing with TCF proteins to prevent transcription of WNT target genes. **(e) FGF signalling.** The secreted FGF ligand binds to tyrosine kinase receptors (FGFR), activating multiple signalling pathways such as the RAS/MAPK, PLC-γ, PI3K and STAT. These pathways culminate in the promotion or repression of FGF target genes by the transcription factor FOS, and the transcription factor families NFAT and FOXO. White bold: genes implicated in classic and MIHV HPE in humans, black bold: genes implicated in classic and MIHV HPE in mice, blue bold: genes identified as modifiers in classic HPE in humans. PrCP: prechordal plate, RVNM: rostral-ventral neural midline.

The model of ventral forebrain patterning that best fits the experimental data is that SHH, expressed in and secreted from the PrCP, signals vertically to the overlying neurectoderm. One consequence of SHH signal reception and transduction by the neurectoderm is that SHH itself becomes expressed in the RVNM, but only if BMP signalling is low in this tissue (Anderson *et al.*, 2002). SHH expression in the RVNM then leads to neuronal patterning and maintenance of ventral forebrain tissue. The anterior mesendoderm of the late gastrula is also the source of WNT antagonism due to the expression of *Dkk1* under the control of the transcription factor *Otx2*. When *Otx2* is conditionally inactivated in the AME, *Dkk1* expression is not activated in the same tissue, and head truncations characteristic of elevated WNT signalling result (Ip *et al.*, 2014). *Shh* expression is not perturbed by the lack of *Dkk1* expression and HPE does not arise, clearly indicating that WNT ligands do not direct morphogenetic activity required for forebrain separation. None-the-less, classic HPE is associated with loss of forebrain tissue (i.e. anterior truncation). It is possible that this is a consequence of loss of midline tissue which, at the anterior, expresses FGF ligands from the anterior neural ridge (ANR) as part of Anterior-Posterior neural patterning. Depletion of midline FGF will cause posteriorisation and forebrain hypoplasia. Thus, anterior truncation often occurs as a secondary consequence of an aberrant ventral pattern. An anterior truncation phenotype can be seen in isolation from ventral defects, however, and should not of itself be used to infer a role in HPE.

1.6 PrCP morphogenesis during murine development

As described above, cells in the PrCP synthesise and secrete SHH, the critical ligand for ventral neural patterning. It follows that classic HPE can also arise if the cells of the PrCP are not formed or do not function properly. The PrCP arises due to a series of inductive interactions between cells at the anterior primitive streak (APS) and then amongst the axial tissues generated by movement of cells through the APS (Figure 1.5B). As described above, the PrCP is formed from cells that pass through the APS early in gastrulation (Kinder *et al.*, 2001). Some cells that transit the APS and migrate anteriorly adopt a paraxial fate (i.e. either side of the midline) and form the anterior definitive endoderm (ADE). By mid gastrulation, the APS cells have been organised into a structure called the node. The cells of the epiblast which transit through the node of the mid-gastrula differentiate into AME and migrate to the anterior midline, taking up a position adjacent to, but posterior of, the cells of the PrCP. These cells form the anterior notochord (ANC) (Kinder *et al.*, 2001). Inductive interactions between these anterior tissues (ADE, PrCP and ANC) influence not only forebrain patterning (Hallonet *et al.*, 2002), but also the fate and survival of the PrCP. If the PrCP is removed, the ANC reconstitutes new PrCP tissue. Conversely, removal of the ANC results in failed PrCP development, indicating that the ANC promotes survival of the PrCP (Camus *et al.*, 2000).

The molecular nature of the signals required for PrCP development can also be inferred by the phenotype of mouse mutants. An extensive series of murine alleles in *Nodal* itself, or components of the NODAL signal transduction pathway (Figure 1.6A), demonstrate that NODAL signalling during gastrulation is required for steps 1 and 2 of PrCP development (Figure 1.5B). Even a small decrease in NODAL signalling activity prevents differentiation of the ADE and PrCP precursors via transit through the APS (Norris *et al.*, 2002; Vincent *et al.*, 2003) and manifests as moderate anterior truncation. Further loss also interferes with node induction by the APS and results in severe anterior truncation, absent AME, and somite fusion across the midline (Chu *et al.*, 2004; Dunn *et al.*, 2004; Episkopou *et al.*, 2001; Liu *et al.*, 2004). These mutations affect processes beyond ventral neural patterning and obscure the role for NODAL in preventing HPE. However, embryos that are heterozygous null for Nodal and null for a closely related TGF-β signalling molecule (Gdf1) develop HPE which arises due to abberant anterior notochord and PrCP development (Andersson *et al.*, 2006) clearly indicating the role for the non-BMP part of the TGF-β pathway in PrCP development. The identity of the signals between the anterior tissues (ADE, PrCP and ANC) that stabilize PrCP fate are unknown, but SHH is a candidate for the survival signal sent from the ANC to the PrCP as indicated by chimeric experiments between $Shh^{-/-}$ and wild type cells (Aoto *et al.*, 2009).

1.7 *Zic2* mutation and ventral forebrain patterning

Severe loss-of-function alleles of murine *Zic2,* such as the *Kumba* (*Ku*) allele, result in classic HPE (Figure 1.7), suggesting an involvement in ventral neural patterning. Gene expression studies rule out the possibility that the ZIC2 transcription factor directly regulates *Shh* expression in either the PrCP or RVNM since neither tissue is a site of *Zic2* expression at the appropriate stage (Elms *et al.*, 2004; Nagai *et al.*, 1997). Similarly, the lack of RVNM expression at the time at which ventral pattern is imposed implies that *Zic2* is not part of the transcriptional response to HH signalling in the murine forebrain. None-the-less, the high similarity between ZIC and GLI zinc finger domains and the finding that ZIC and GLI proteins can physically interact via their ZFDs (Kinzler and Vogelstein, 1990; Mizugishi *et al.*, 2001; Pavletich and Pabo, 1993) suggested ZIC2 could act downstream of SHH signalling in the forebrain (Roessler and Muenke, 2001). This hypothesis was directly tested by cross of the *Ku* allele of *Zic2* ($Zic2^{Ku}$ MGI:106679) with the *Shh* null allele (Shh^{tm1Chg} MGI: 1857796) (Warr *et al.*, 2008). The *Ku* mutant carries a missense mutation in the 4th zinc finger that abolishes the DNA binding and transcriptional activation ability of ZIC2 (Elms et al. 2003, Brown et al, 2005). When intercrossed, it was observed that neither gene was sensitive to a decreased dose of the other, and double homozygous embryos exhibited a novel phenotype demonstrating that ZIC2 does not act downstream of SHH in murine forebrain development (Warr *et al.*, 2008). Moreover, the same study

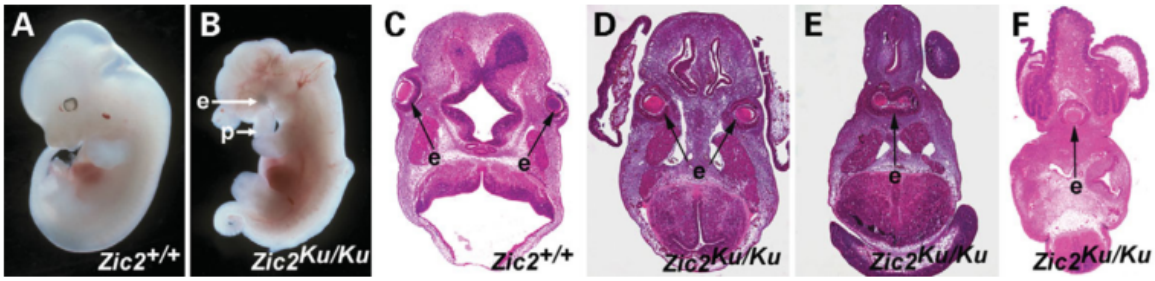


Figure 1.7: The *Kumba* allele of *Zic2*-associated HPE. Compared to **(a)** wildtype (*Zic2*$^{+/+}$) embryos at 12.5 dpc, **(b)** *Kumba* (*Zic2*$^{Ku/Ku}$) embryos exhibit exencephaly, spina bifida, a looped tail, an internally located eye and a proboscis. When sectioned transversely, **(d)** internal and close spaced eyes (hypotelorism), **(e)** internal and incompletely separated eyes (synophthalmia) and **(f)** cyclopia can all be seen in *Kumba* embryos, in contrast to **(c)** a wildtype embryo with two externally located eyes and divided left and right forebrain hemispheres. e: eye, p: proboscis. Figure is modified from (Warr *et al.*, 2008).

showed that a phenotype was present in *Zic2*$^{Ku/Ku}$ embryos before the stage at which *Shh* expression is first detected, and that germline loss of all HH signalling (via *Smo* deletion) does not reproduce the early aspects of the *Zic2*$^{Ku/Ku}$ phenotype (Warr *et al.*, 2008).

Instead, it appears that ZIC2 intersects the NODAL signalling pathway at mid gastrulation. *Nodal* loss-of-function is lethal at gastrulation and compound heterozygous embryos for both ZIC2 and NODAL do not survive to the forebrain stage of development. Sequentially decreasing the dose of NODAL activity on the *Zic2*$^{Ku/Ku}$ background shifts the *Nodal* phenotype towards the more severe end of the spectrum (increased frequency and severity of anterior truncation) (Houtmeyers *et al.*, 2016). Evidently, in the absence of ZIC2 function, the embryos perceive a lower dose of NODAL signalling, suggesting that *Zic2* normally promotes NODAL signalling at the APS (Figure 1.5B). This is supported by the analysis of *Zic2*$^{Ku/Ku}$ embryos which show that in the absence of *Zic2* function, the derivatives of the APS (i.e. the ADE and PrCP cells) are specified and migrate to the embryonic anterior to take up their normal position (Warr *et al.*, 2008) and the node is induced (Elms *et al.*, 2003). However, gene expression at the newly induced node is highly aberrant; the expression of every node specific gene so far examined at mid-gastrulation in *Zic2*$^{Ku/Ku}$ embryos is depleted (Barratt *et al.*, 2014; Warr *et al.*, 2008). Cell death and proliferation of the ANC cells that emerge from the mid-gastrula node is unaltered, but the transcripts of genes that mark the emerging ANC are depleted, suggesting that this tissue is not specified. Despite the earlier evidence of PrCP formation, by late gastrulation the expression of markers characteristic of the PrCP is absent in *Ku* embryos. Consequently, *Shh* expression in the PrCP is not activated and the expression of *Shh* and SHH target genes in the RVNM is not initiated (Warr *et al.*, 2008).

The analysis of the *Zic2*$^{Ku/Ku}$ phenotype suggests that PrCP development fails at stage 3, and that the earliest identified molecular and functional abnormalities are at the mid-gastrula node (Figure 1.5B, C). This is therefore considered the stage and site of primary *Zic2* function (Warr et al, 2008). This functional analysis is consistent with the node of the mid-gastrula embryo (the structure that produces the ANC) being the only unique site of *Zic2* gene expression at this stage of development compared to other ZIC family members. Other closely related *Zic* genes (*Zic3* and *Zic5*) are co-expressed with *Zic2* in all other areas of the gastrula at this stage, and likely compensate for ZIC2 loss-of-function in these cells (Elms *et al.*, 2004; Furushima *et al.*, 2000). The precise molecular role of ZIC2 at the mid-gastrula node remains unclear. The level of *Nodal* transcript is unaltered in *Zic2*$^{Ku/Ku}$ embryos, indicating that ZIC2 does not promote NODAL activity by directly controlling *Nodal* expression, but instead acts downstream of the NODAL signal (Houtmeyers 2016). Another hypothesis is that ZIC2 directly regulates expression of the *Foxa2* transcription factor. *Foxa2,* a NODAL target gene (Hoodless *et al.*, 2001) expressed in the APS,

node and AME (Ang *et al.*, 1993; Ang and Rossant, 1994; Dufort *et al.*, 1998; Monaghan *et al.*, 1993; Ruiz i Altaba *et al.*, 1993; Sasaki and Hogan, 1993) is known to control *Shh* expression (Jeong and Epstein, 2003). In turn, SHH can induce *Foxa2* expression (Echelard *et al.*, 1993). A scenario in which, during normal development, ZIC2 controls *Foxa2* expression in the node and its derivative ANC cells to initiate the *Foxa2*/*Shh* auto-induction loop and eventually provide the SHH-based survival signal to stabilise PrCP cell development is consistent with the phenotype analysis of the $Zic2^{Ku/Ku}$ embryos.

When overexpressed in mammalian cell lines, ZIC2 is able to physically interact with both SMAD2 and SMAD3 (the receptor activated proteins that control transcription in a NODAL dependent manner) (Figure 1.5A). When bound to SMAD proteins, ZIC2 opposes SMAD activity (it dampens SMAD dependent transcription or overcomes SMAD dependent repression). In cultured human cells, ZIC2 can act in concert with SMAD3 to promote *FOXA2* expression, but the ZIC2 protein encoded by the *Ku* allele of ZIC2 is unable to do so, despite still physically interacting with SMAD (Houtmeyers *et al.*, 2016). Overall, the cell based data, in combination with the genetic evidence that ZIC2 is required to promote NODAL signalling, supports a model in which expression of node specific enhancers is initially repressed and subsequently converted to expression activation in the presence of SMAD/ZIC2 complexes. The proposed molecular interactions between ZIC2 and SMAD molecules, and between this complex and SMAD DNA binding elements, are yet to be demonstrated *in vitro*.

1.8 Dorsal forebrain patterning during murine development

Initial dorsal forebrain patterning is a consequence of inductive interactions between the anterior neurectoderm and the adjacent surface ectoderm (Furuta *et al.*, 1997; Liem *et al.*, 1995). Both of these tissues arise during gastrulation (for review see Arkell and Tam 2012) with the origin of the anterior neurectoderm already described above. Prior to gastrulation, precursor cells for the surface ectoderm are found at the future anterior side of the embryo, about half way down the embryonic portion of the epiblast (Figure 1.5A). Like the prospective neurectoderm, these cells do not pass through the primitive streak during gastrulation but differentiate *in situ* into surface ectoderm. During gastrulation, the cells are arranged an anterior-posterior order, but in a more lateral position than the neurectoderm cells. The progenitors of another non-neural ectoderm derivative, the neural crest cells (which gives rise to the ecto-mesenchyme and cranial ganglia in the head), are juxtaposed between the neurectoderm and surface ectoderm cells at a region known as the neural plate border (Figure 1.5B, 7.5 dpc). The entire ectoderm arises as a contiguous sheet which, in a process called neurulation, folds to form the neural tube and overlying surface ectoderm. Consequently, cells from the medial neural plate take up a ventral position and cells from the lateral neural plate

adopt a dorsal position. Once the process of neurulation is complete, the forebrain neurectoderm at the dorsal midline continues its morphogenesis and, by the combined strategies of low mitosis and high apoptosis, undergoes thinning and invagination to divide the cerebrum into two hemispheres (Figure 1.8A) (Furuta *et al.*, 1997; Groves and LaBonne, 2014). In a secondary phase of dorsal patterning, inductive interactions between the invaginated cells instruct the differentiation of specialised midline dorsal structures. These include the choroid plexus which will secrete cerebrospinal fluid, and the cortical hem which instructs adjacent neurectoderm to differentiate into the hippocampus (Groves and LaBonne, 2014; Hébert *et al.*, 2002).

Much of our knowledge regarding dorsal patterning of the neurectoderm comes mainly from studies of the spinal cord (reviewed in Le Dréau and Martí 2012), but aspects of forebrain neurectoderm patterning involve distinct mechanisms. For example, in the forebrain, dorsal patterning occurs in close proximity to the ANR, a source of morphogenetic signals which sits at the rostral junction of the neural and surface ectoderm (Crossley and Martin, 1995; Shimamura and Rubenstein, 1997). The establishment of forebrain dorsal-ventral pattern therefore intersects and interacts with the anterior-posterior patterning system. Additionally, the forebrain neurectoderm arises earliest of all the neurectoderm subdivisions and dorsally restricted gene expression patterns (such as that of *Pax3* [Goulding et al. 1991]) are evident from the time of somite formation. Despite this, the first overt signs of dorsal differentiation in the forebrain (i.e. roof-plate thinning) occur in embryos with approximately 20 somites. In contrast, dorsal development is evident in the hindbrain 24 hours earlier when the neural crest first emerges from the dorsal neurectoderm of the 5-somite embryo (Serbedzija *et al.*, 1990). The influences that promote dorsal differentiation in the hindbrain region must operate in concert with, or very soon after, the signals late in gastrulation that induce the anterior epiblast to adopt a neural fate. Presumably dorsal-ventral patterning of the forebrain also begins during gastrulation and, as in other A-P regions, the early forebrain neurectoderm has a dorsal character which subsequently undergoes modification to direct the secondary differentiation events of midline development. Given the timing of dorsal differentiation, it is possible that the neural crest derived mesenchyme overlying the dorsal forebrain provides instructive signals to drive invagination of the telencephalic midline (Choe *et al.*, 2014), similar to the role of retinoic acid in the chick (Gupta and Sen, 2016). There remains much to learn about the timing and source of the various morphogenetic signals that control dorsal forebrain fate and a proposed model of this multi-step process is shown in Figure 1.8A.

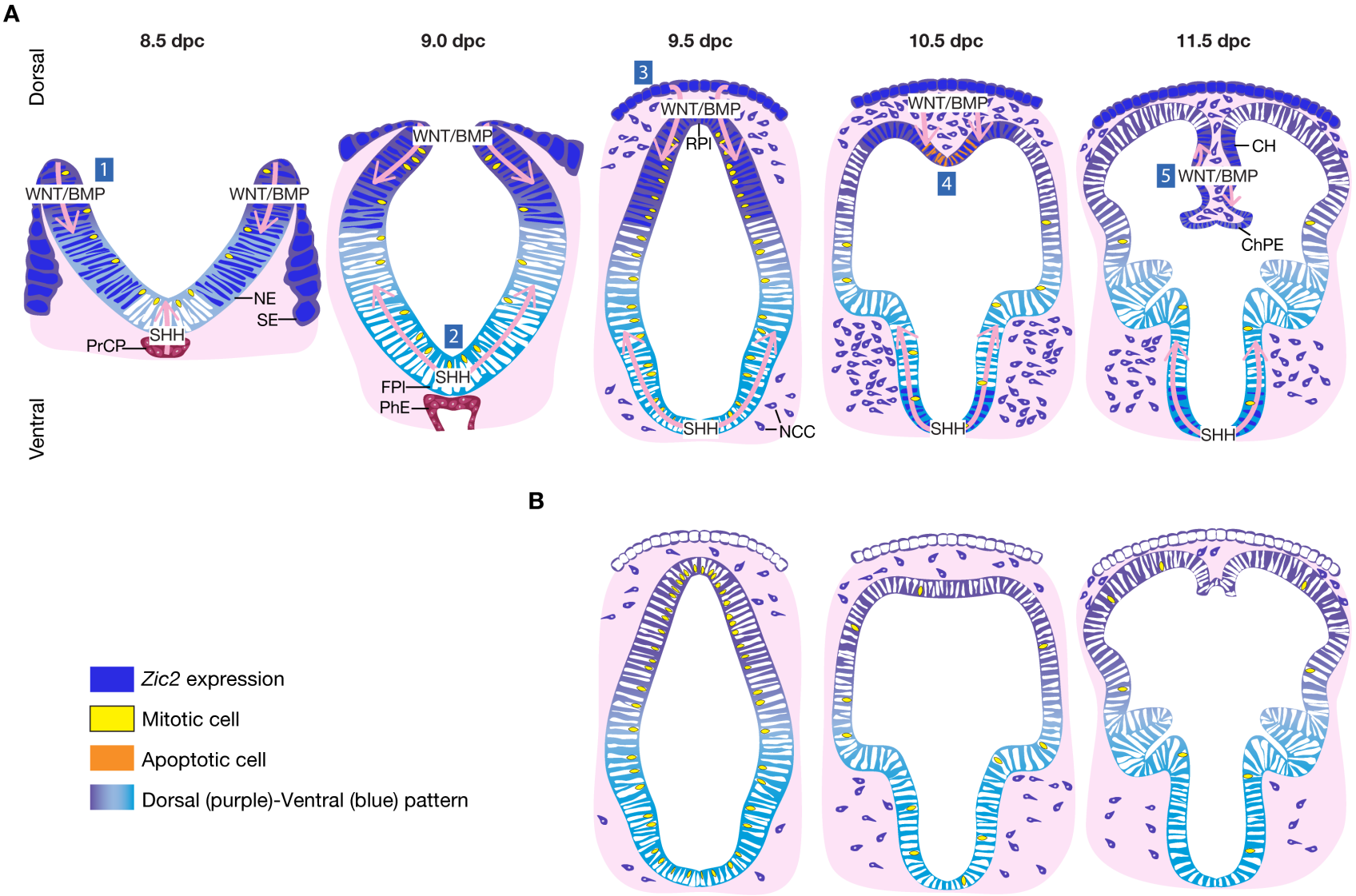

A
8.5 dpc
9.0 dpc
9.5 dpc
10.5 dpc
11.5 dpc
Dorsal
Ventral
1
WNT/BMP
WNT/BMP
NE
SE
SHH
PrCP
2
WNT/BMP
SHH
FPl
PhE
3
WNT/BMP
RPl
SHH
NCC
4
WNT/BMP
SHH
5
WNT/BMP
CH
ChPE
SHH
B
Zic2 expression
Mitotic cell
Apoptotic cell
Dorsal (purple)-Ventral (blue) pattern

Figure 1.8: Proposed model of dorsal telencephalic midline development and *Zic2*-associated MIHV HPE. (a) Diagrams of transverse sections of the telencephalon at 8.5 – 11.5 dpc. **Step 1:** due to the influence of WNT and BMP signals from the border region, the forebrain neurectoderm is initially dorsal in character, as shown by *Zic2* expression throughout most of the neurectoderm. SHH signals ventrally from the PrCP to the overlying neurectoderm. **Step 2:** *Shh* expression is established in the rostral-ventral neural midline, which overlays ventral identity information onto the neurectoderm. WNT and BMP signalling, and *Zic2* expression become restricted to the dorsal half of the neurectoderm. **Step 3:** once neurulation is complete and the surface ectoderm overlies the neurectoderm, neural crest cells migrate in from the hindbrain and surround the dorsal neurectoderm. The dorsal-most neurectoderm cells exit the cell cycle. **Step 4:** The neural crest cells expand under the influence of WNT signalling and the dorsal neurectoderm cells undergo apoptosis, leading to thinning and invagination of the roof-plate. *Zic2* expression initiates in the ventral midline. **Step 5:** BMP and WNT signalling in the invaginated tissues induce the cortical hem, choroid plexus epithelium and the hippocampus. **(b)** In $Zic2^{kd/kd}$ embryos at 9.0 - 9.5 dpc, WNT expression in the surface ectoderm is delayed. Consequently, dorsal neurectoderm cells do not exit the cell cycle and neural crest infiltration is reduced. At 10.5 dpc the dorsal cells do not undergo apoptosis and invagination of the roof-plate does not occur. Though there are few neural crest cells, WNT expression is initiated. By 11.5 dpc, dorsal midline structures are absent or hypoplastic, resulting in MIHV. PrCP: prechordal plate, NE: neurectoderm, SE: surface ectoderm, FPl: floor-plate, PhE: pharangeal endoderm, RPl: roof-plate, NCC: neural crest cells, CH: cortical hem, ChPE: choroid plexus epithelium.

The identity of the molecular signals that instruct dorsal patterning has been elusive, perhaps because of redundancy or the iterative use of the same signalling pathway during dorsal patterning and midline structure differentiation (Figure 1.8). Members of the BMP and WNT signalling molecule families are expressed in a spatial-temporal manner consistent with a role in dorsal neural patterning. Five BMP ligands (BMP2, BMP4-7) are expressed at the future dorsal midline in the forebrain neurectoderm and surface ectoderm. Embryos null for either *Bmp4* or *Bmp2* die before neurulation is complete (Winnier *et al.*, 1995; Zhang and Bradley, 1996), whilst *Bmp4* conditional mutants develop a phenotypically normal telencephalon (Hébert *et al.*, 2003). Similarly, no neural phenotype is seen in *Bmp5, 6* and *7* mutants (Dudley *et al.*, 1995; Kingsley *et al.*, 1992; Luo *et al.*, 1995; Solloway *et al.*, 1998). A role in the relatively late events of dorsal forebrain patterning is, however, revealed when the BMP-specific receptors are mutated in a time and tissue dependent manner. Animals that are constitutive null for *Bmpr1b* and in which *Bmpr1a* is deleted in the telencephalon at ~ 9.0 dpc do not show the characteristic thinning of the roof-plate at 10.5 dpc that is required for dorsal hemisphere separation. Subsequently, they exhibit MIHV HPE and loss of all dorsal midline cell types (i.e. the choroid plexus and cortical hem fail to form) despite maintenance of *Zic2* expression. The specification of ventral and cortical cell types, however, remain unaffected (Fernandes *et al.*, 2007). Once the midline cells invaginate, BMP expression is initiated in the choroid plexus epithelium anlagen (Currle *et al.*, 2005) and overexpression of a constitutively active BMPR1a transforms cortical precursors into choroid plexus cells (Panchision *et al.*, 2001), suggesting BMP signalling induces choroid plexus cell fate. When the roof-plate is ablated prior to differentiation of the choriod plexus and cortical hem, these structures fail to form. The expression of some dorsal midline genes (but not *Zic2*) can be rescued in tissue explants from roof-plate-ablated embryos, however, by culture in BMP4 (Cheng *et al.*, 2006). These experiments, therefore, establish a primary role for BMP signalling in the prevention of MIHV HPE.

Similarly, *Wnt1* and *Wnt3a* are expressed in the future dorsal neurectoderm along the length of the axis prior to neural tube closure, and in the roof-plate following closure (Megason and McMahon, 2002; Parr *et al.*, 1993). In the canonical WNT signalling pathway (Figure 1.6D), binding of WNT ligand to a cognate receptor complex stimulates a cascade of cytoplasmic events culminating in β-catenin nuclear entry. Nuclear β-catenin associates with transcription factors of the TCF/LEF family and converts target gene repression to activation (reviewed in Arkell et al. 2013). WNTs were primarily considered mitogenic signals for the neurectoderm until elevated WNT signalling was shown to alter progenitor gene expression along the dorsal-ventral axis, promoting the production of dorsal progenitors and suppressing ventral progenitors. WNT signalling acts via TCF/LEF dependent evolutionarily conserved enhancers to establish the dorsal domain of *Gli3* expression (Yu *et al.*, 2008). In turn, GLI3, acting as a transcriptional repressor,

inhibits the ventrally produced SHH signal (Alvarez-Medina *et al.*, 2007). This interplay of SHH and WNT signalling may also be relevant for telencephalon patterning as evidenced by the effect of differential regulation of GLI3 upon the ventral and dorsal telencephalic neuronal subtypes generated from human embryonic stem cells (Li *et al.*, 2009). A further role for WNT signalling at later stages of roof-plate development is revealed by conditional deletion of β-catenin in the neural crest cells abutting the telencephalic neurectoderm which causes a failure of neural crest cell expansion and of telencephalic midline invagination (Choe *et al.*, 2014).

In the forebrain, the dorsal-ventral morphogenetic signals of SHH, BMP and WNT intersect those provide by the ANR; a source of FGF ligands (FGF8, FGF15/19, FGF17 and FGF18). When FGF ligands bind to tyrosine kinase receptors (FGFR), multiple signalling pathways are activated such as the RAS/MAPK, PLC-γ, PI3K and STAT pathways (Dailey *et al.*, 2005) (Figure 1.6E). When FGF signalling in the telencephalon is attenuated or abolished via mutation of *Fgf8,* the FGF receptor *Fgfr1* alone, or other receptor combinations, mice exhibit telencephalic hypoplasia. In embryos lacking *Fgfr1* and *Fgfr2* in the telencephalon, the decreased size of the forebrain is attributed to reduced cell proliferation in the ventral midline along with increased apoptosis in the dorsal midline (Gutin *et al.*, 2006; Storm *et al.*, 2006). Defects in these animals resemble those defects associated with the ventral forms of HPE. As reviewed by Hoch et al. (2009), there is a complex series of interdependencies between the forebrain signalling centres. For example, the ventral SHH signal maintains rostral midline FGF ligand expression (Hayhurst *et al.*, 2008; Ohkubo *et al.*, 2002) and simultaneously, the dorsal neurectoderm expression of the repressor form of GLI3 (GLI3R) represses FGF ligand dorsally (Theil *et al.*, 1999). Similarly, the rostral source of FGF works via the *Foxg1* transcription factor (Gutin *et al.*, 2006; Hébert and Fishell, 2008) to inhibit BMP signalling, thus maintaining rostral proliferation and preventing premature differentiation of neuronal progenitor cells (Dou *et al.*, 2000, 1999; Hanashima *et al.*, 2004; Shimamura and Rubenstein, 1997). Human genetics indicates that anterior FGF signalling impacts dorsal forebrain patterning (Dubourg *et al.*, 2016) but, despite evidence that FGF signalling can control *Zic2* expression in the 9.5 dpc telencephalon (Hayhurst *et al.*, 2008; Okada *et al.*, 2008), it remains unclear whether FGF control of *Zic2* expression plays a role in *Zic2*-associated MIHV.

1.9 *Zic2* mutation and dorsal forebrain patterning

Partial loss-of-function alleles of murine *Zic2* result in MIHV HPE, suggesting an involvement for ZIC2 protein in dorsal neural patterning and differentiation (Nagai et al. 2000). Late in gastrulation, as the neural plate is forming, *Zic2* expression recedes from the posterior embryo proper and becomes restricted to the anterior neuroectoderm. By this time, *Zic3* and *Zic5* expression is also restricted to this region. The expression of each of these genes then subsides in the medial neural plate, becoming progressively confined to the lateral (future dorsal)

neuroectoderm and the flanking surface ectoderm (Elms *et al.*, 2004; Houtmeyers *et al.*, 2013). After the neural tube closes, high levels of *Zic2* expression can be detected along the entire anterior-posterior extent of the neural tube including the dorsal telencephalon (roof-plate and hippocampal primordium) (Cheng *et al.*, 2006; Okada *et al.*, 2008) with extensive overlap in this expression domain of all Zic family members. *Zic2* is therefore expressed in a manner consistent with a role in dorsal patterning and indeed, loss-of-function mutations in *Zic2* lead to defects in neural crest development, a dorsal cell type of the hindbrain and spinal cord (Elms *et al.*, 2003; Nagai *et al.*, 2000).

Embryos homozygous for a hypomorphic allele of *Zic2* in which reduced amounts of the normal transcript are produced (Nagai *et al.*, 2000) lack a telencephalic roof-plate at mid-gestation. After neural tube closure in wild type embryos, the dorsal midline of the telencephalon immediately becomes devoid of mitotic cells and, within 24 hours, apoptosis in this tissue is noticeably higher than in the surrounding tissues (Figure 1.8A). $Zic2^{kd/kd}$ embryos exhibit neither of these features and, consequently, roof-plate thinning and invagination does not occur. Subsequently, the structures that should be derived from the dorsal midline are either severely hypoplastic or absent (Figure 1.8B). At this stage of development, *Wnt3a* should be expressed in the dorsal midline of the forebrain and along the spinal cord until the position of the forelimb bud. In $Zic2^{kd/kd}$ embryos this expression is delayed such that it has only just been initiated in the dorsal forebrain (Nagai et al. 2000). This work firmly connects ZIC2 function to the MIHV form of HPE, but many questions remain unanswered regarding the role of ZIC2 in roof-plate formation. For example, it is not known precisely when the process of dorsal patterning and roof-plate induction fails in $Zic2^{kd/kd}$ embryos. This may reflect a role for ZIC2 in earlier dorsal patterning events rather than roof-plate induction, per se. It is also not clear whether the expression of *Zic2* in the neurectoderm, flanking surface ectoderm or both tissues is required for roof-plate formation. Alternatively, it is possible that the documented role of ZIC2 in neural crest cell development is important since in $Zic2^{kd/kd}$ embryos there is an evident lack of neural crest cells in the intrahemispheric mesenchyme. Furthermore, ZIC2 is known to be able to physically interact with transcriptional mediators of the WNT, TGF-β and SHH pathways (Fujimi *et al.*, 2012; Houtmeyers *et al.*, 2016; Koyabu *et al.*, 2001; Pourebrahim *et al.*, 2011) and whether it does so during dorsal patterning is unknown. The experiments on $Zic2^{kd/kd}$ embryos highlight that roof-plate formation and dorsal midline development are particularly sensitive to loss of ZIC2 levels, with a reduction of *Zic2* expression to ~20% sufficient to generate MIHV HPE in mice (Nagai *et al.*, 2000).

1.10 The dual role of *Zic2* in HPE

Current knowledge indicates that ZIC2 plays two distinct roles in forebrain development. At mid-gastrulation, ZIC2 functions at the node to shape the NODAL gradient. Severe Zic2 loss-of-function mutations lead to a transient failure in the production of the ANC and, consequently, PrCP fate is not stabilised. As such, the SHH signal is not sent from the PrCP and the ventral forebrain signalling centre is not established, resulting in gross perturbations in forebrain D–V pattern and classic HPE. This requirement appears well buffered, as mild loss-of-function is compatible with correct function of the ventral forebrain signalling centre in man and mouse. A second requirement for ZIC2 function occurs in the dorsal neurectoderm of the developing telencephalon where it acts to promote formation of the dorsal signalling centre responsible for roof-plate and choroid plexus development. It is clear that a failure of ZIC2 function in the developing dorsal forebrain results in MIHV, but just how ZIC2 intersects the interconnected forebrain signalling network is unknown. This requirement appears more sensitive to ZIC2 levels than ventral forebrain patterning, as hypomorphic mutations have been associated with MIHV in both man and mouse (Brown *et al.*, 2005, 2001; Nagai *et al.*, 2000; Solomon *et al.*, 2010a). The precise molecular role of ZIC2 in both classic and MIHV HPE, and how *ZIC2* expression is regulated in the gastrulating embryo, however, remain a subject of ongoing investigation.

1.11 Non-coding conserved elements at the *ZIC2* locus

In contrast to the coding genome, the size of the non-coding genome scales with organ complexity (Levine and Tjian, 2003), and researchers have begun to elucidate the role of the non-coding genome in regulating gene expression. Comparative genomics can be used to identify coding regions that may be important for the control of gene expression. Since DNA that is non-essential for an organisms survival accrues nucleotide variation such as insertions, deletions and SNVs, the relatively slow evolving (or conserved) sequences within the genome can be catalogued via cross-species genome comparisons (Katzman *et al.*, 2007; Vavouri *et al.*, 2007; Vavouri and Lehner, 2009). These sequences have been termed non-coding conserved elements (NCEs), with studies showing NCEs are frequently more conserved than protein coding genes (Bejerano *et al.*, 2004; Woolfe *et al.*, 2005) and are 300 times less likely to be lost in evolution than neutrally evolving DNA (McLean and Bejerano, 2008). The compact genome of the pufferfish (*Takifugu rubripes*), who last shared a common ancestor with humans 450 million years ago, contains 1,373 NCEs that show 84% similarity to human and mouse NCEs (Woolfe *et al.*, 2005). Up to 92% of these conserved sequences exhibit enhancer activity in zebrafish reporter assays, and 50% exhibit activity in mouse reporter assays (Visel *et al.*, 2008; Woolfe *et al.*, 2005). A large portion of NCEs identified so far are associated with transcription factors and signalling molecules essential for embryo development (Bejerano *et al.*, 2004; Woolfe *et al.*, 2005). Similar to developmental genes, both vertebrate and invertebrate NCEs tend to cluster

together in genomic regions on chromosomes (Bejerano *et al.*, 2004; Vavouri *et al.*, 2007; Woolfe *et al.*, 2005), with their target genes commonly within a few megabases distance (Vavouri *et al.*, 2006). ZIC1, ZIC2 and ZIC4 have all previously been associated with such clusters of NCEs in vertebrates, nematodes and arthropods (Vavouri *et al.*, 2007).

In 2012, Roessler *et al.* at the National Human Genome Research Institute (NHGRI), National Institute of Health (NIH) used EvoPrint and ECRbase to screen for NCEs in *ZIC2* and the surrounding 10 kb of DNA, identifying an 804 bp conserved structure within the *ZIC2* 3'UTR (Figure 1.9) (Roessler *et al.*, 2012a). Furthermore, the study identified six single nucleotide variants (SNVs) in the NCE of seven HPE probands (Figure 1.9, Table 1.5), none of which were found in control populations, indicating the region was under selective pressure. It was therefore hypothesised that this NCE regulates *ZIC2* expression during embryogenesis and prevents HPE development. The exact nature of this regulation, however, remains to be determined, and could occur at either the transcription or the post-transcription stages of gene expression. The NCE could act as an enhancer/repressor element (ER element), regulating transcription of a target gene via the binding of protein complexes to transcription factor binding sites (TFBSs) within the NCE. The NCE may also transcribe non-coding RNAs (ncRNAs) such as MicroRNAs (miRNA), Long Non-Coding RNAs (lncRNA) or Small Interfering RNAs (siRNA; Kaikkonen *et al.*, 2011). Finally, as the NCE is situated within the 3'UTR, it could direct mRNA transcript stability and decay, transcript polyadenylation, or mRNA localisation and translation (Barrett *et al.*, 2012). It is likely that a combination of these processes are to regulate *ZIC2* transcription during embryogenesis.

1.12 Scope of this Thesis

This study used a combination of mouse genetics and heterologous reporter assays to investigate two unexplored areas of *ZIC2*-associated HPE:

i. Do the node defects which underlie *Zic2*-associated classic HPE also impact the establishment of the left-right axis at gastrulation, resulting in co-morbidities such as cardiac defects?
ii. Are the six SNVs in the *ZIC2* 3'UTR likely to be pathogenic?

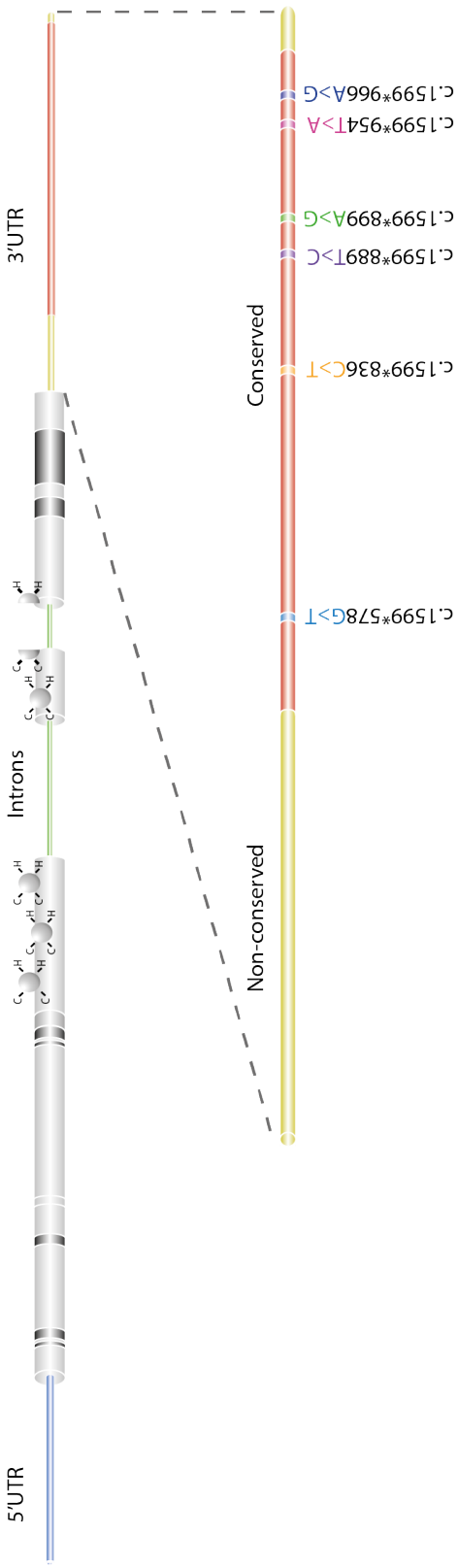


Figure 1.9: The *ZIC2* 3'UTR and NCE. Roessler et al. identified an 804 bp Non-Coding Conserved Element (NCE) (red) in the *ZIC2* 3'UTR (yellow). Genetic screens of human HPE patients identified six SNVs in this NCE region (shown) and two polymorphisms (not shown).

Table 1.5: The six human *ZIC2* 3'UTR mutations and two polymorphisms. This table is modified from (Roessler *et al.*, 2012a). MAF: minor allele frequency (%) in a HPE cohort and a control cohort. Variants were identified in a cohort of 528 HPE patients and compared to a control population of ≥372 individuals of mixed ethnicity.

a – common variant in healthy individuals

b- affected sibling of proband LCL301.

c – in cis with enhancer variant by co-amplification and sequencing

Name	**Variant**	**Subject**	**Inheritance**	***ZIC2* coding mutations**	**Other HPE mutations**	**HPE type**	**MAF (HPE)**	**MAF (control)**
Poly	c.1599*456G>A	Rs13542 dbSNP	N.A.	N.A.	N.A.	Unknown	NA	33.00
M1	c.1599*578T>A	LCL1349	Mat negative Pat N.A.	-	SHH - c.419A>C p.His140Pro (Mat)	Microform	0.095	0.00
Poly	c.1599*587G>T	FB9622 LCL7282 LCL6386	N.A. N.A. N.A.	- - -	TGIF – c.487C>T p.Pro163Ser[a] - -	Unknown	0.280	0.66
M2	c.1599*836C>T	Brz2172	N.A.	c.1215dupC p.Ser406Glnfs*11[c]	N.A.	Unknown	0.095	0.00
M3	c.1599*889T>C	AM6632	N.A.	-	TGIF – c.420A>G p.Pro140Pro TGIF – c.488C>T p.Prop163Leu[a]	Alobar	0.095	0.00
M4	c.1599*899A>G	LCL301 LCL7897[b]	Pat positive	c.1059C>T p.His353His	SHH - c.72C>A p.cys24*(Pat)	Microform	0.190	0.00
M5	c.1599*954T>A	Brz37	N.A.	N.A.	N.A.	Unknown	0.095	0.00
M6	c.1599*966A>G	LCL7828	Mat positive Pat N.A.	c.1059C>T p.His353His	-	Semilobar	0.095	0.00

Chapter 2: Methods

2.1 Generation of Constructs

The complete list of constructs used in this thesis can be found in Appendix Table A1.1. The following constructs were generated by myself, unless otherwise stated:

2.1.1 Expression Constructs

pENTR3C-*Foxa2* – The 1380 bp murine *Foxa2* coding DNA sequence (CDS) was amplified from pBSK-*Foxa2* (Sasaki and Hogan, 1993) using oligonucleotides containing *Xho*I and *Sal*I restriction sites (RA1825 and RA1824; Appendix Table A1.2). Following ammonium acetate clean up (Section 2.2.10), the *Foxa2* fragment and Gateway Entry vector pENTR3C were digested with *Xho*I and *Sal*I, releasing the *ccdB* gene fragment from pENTR3C. The *Foxa2* fragment was ligated into pENTR3C via traditional T4 ligation methods. In the subsequent pENTR3C-*Foxa2* construct, *att*L sites flank the *Foxa2* CDS for Gateway LR based cloning.

pV5-*Foxa2* – pV5-*Foxa2* was generated via Gateway LR cloning (Section 2.2.14) of the pcDNA3.1-nV5-DEST (Invitrogen) Destination vector and pENTR3C-*Foxa2*. The *Foxa2* fragment was inserted into pcDNA3.1-nV5-DEST 3' to the V5 epitope tag.

pENTR1A-*Foxj1* – The 1266 bp murine *Foxj1* CDS was amplified from pYX-ASC-*Foxj1* (Cruz *et al.*, 2010) using oligonucleotides containing *Kpn*I and *Not*I restriction sites (RA1545 and RA1546; Appendix Table A1.2). Following ammonium acetate clean up, the *Foxj1* fragment and Gateway Entry vector pENTR1A were digested with *Kpn*I and *Not*I, releasing the *ccdB* fragment from pENTR1A. The *Foxj1* fragment was ligated into pENTR1A via traditional T4 ligation methods. In the subsequent pENTR1A-*Foxj1* construct, *att*L sites flank the *Foxj1* CDS for Gateway LR based cloning.

pV5-*Foxj1* – pV5-*Foxj1* was generated via Gateway LR cloning of the pcDNA3.1-nV5-DEST (Invitrogen) Destination vector and pENTR1A-*Foxj1*. The *Foxj1* fragment was inserted into pcDNA3.1-nV5-DEST 3' to the V5 epitope tag.

pENTR2B-*NOGGIN* – The 699 bp human *NOGGIN* CDS (Accn. No. NM_005450) was amplified from HEK293T genomic DNA via oligonucleotides containing *Kpn*I and *Not*I restriction sites (RA1776 and RA1777; Appendix Table A1.2). Following ammonium acetate clean up, the *NOGGIN* fragment and Gateway Entry vector pENTR2B were digested with *Kpn*I and *Not*I, releasing the *ccdB* fragment from pENTR2B. The *NOGGIN* fragment was ligated into pENTR2B

via traditional T4 ligation methods. In the subsequent pENTR2B-*NOGGIN* construct, *att*L sites flank the *NOGGIN* CDS for Gateway LR based cloning.

pV5-*NOGGIN* – pV5-*NOGGIN* was generated via Gateway LR cloning of the pcDNA3.1-nV5-DEST (Invitrogen) Destination vector and pENTR2B-*NOGGIN*. The *NOGGIN* fragment was inserted into pcDNA3.1-nV5-DEST 3' to the V5 epitope tag.

pENTR3C-*ZIC2* – The 1599 bp human *ZIC2* CDS was cloned into pENTR3C by Jerry Ahmed (Arkell laboratory). In the subsequent pENTR3C-*ZIC2* construct, *att*L sites flank the *ZIC2* CDS for Gateway LR based cloning.

pV5-*ZIC2* – pV5-*ZIC2* was generated by Radiya Ali (Arkell laboratory) via Gateway LR cloning of the pcDNA3.1-nV5-DEST (Invitrogen) Destination vector and pENTR3C-*ZIC2*. The *ZIC2* fragment was inserted into pcDNA3.1-nV5-DEST 3' to the V5 epitope tag.

pGEM-3'UTR-WT – the 1086 bp wildtype *ZIC2* 3'UTR was amplified from HEK293T genomic DNA via oligonucleotides containing *Xho*I restriction sites (RA1596 and RA1597; Appendix Table A1.2) and TA cloned into pGEM-T-Easy using the pGEM-T-Easy Vector System (Cat. No. A1360, Promega).

pV5-*ZIC2*-3'UTR – pGEM-3'UTR-WT and pV5-*ZIC2* were digested with *Xho*I to release the *ZIC2* 3'UTR fragment and linearize pV5-*ZIC2*. The *ZIC2* 3'UTR fragment was gel extracted and purified, whilst pV5-*ZIC2* was treated with Antarctic Phosphatase to remove the 5' phosphate and minimise self-ligation, and purified via ammonium acetate clean up. The products were ligated via traditional T4 methods and the final pV5-*ZIC2*-3'UTR construct isolated via bacterial colony polymerase chain reaction (PCR). This construct was created with the help of Kathryn Dickson (Arkell laboratory).

2.1.2 Luciferase Assay Constructs

pGL4.20-DEST-*luc* – pGL4.20-*luc* was converted into a Destination vector via the Gateway® Vector Conversion System (Invitrogen). pGL4.20-*luc* was digested with *Eco*RV to linearize the vector and purified via ammonium acetate clean up. It was then taken through the Gateway® Vector Conversion System to ligate the Destination Reading Cassette A into the pGL4.20-*luc* multiple cloning site (MCS).

pGL-β-globin-DEST-*luc* - pGL4.20-DEST-*luc* was digested with *Hind*III and *Bgl*II to linearize the vector and purified via ammonium acetate clean up. The β-globin minimal promoter was PCR amplified from pKS:β-globin:*lacZ* with oligonucleotides containing *Hind*III and *Bgl*II restriction

sites (RA1506 and RA1507; Appendix Table A1.2). The amplicon was subsequently digested and ligated into the linearised pGL4.20-DEST-*luc* via traditional ligation methods.

pGEM-T-Easy-Hsp68 - The heat shock protein 68 (Hsp68) minimal promoter was PCR amplified from pBS-Hsp68 with oligonucleotides containing *Hind*III and *Bgl*II restriction sites (RA1640 and RA1626; Appendix Table A1.2). The amplicon was TA cloned into pGEM-T-Easy to create pGEM-T-Easy-Hsp68.

pGL-Hsp68-DEST-*luc*- pGL4.20-DEST-*luc* and pGEM-T-Easy-Hsp68 were digested with *Hind*III and *Bgl*II to linearize the vector and release the Hsp68 minimal promoter, respectively. The Hsp68 minimal promoter was gel extracted and purified, and pGL4.20-DEST-*luc* purified via ammonium acetate clean up. The Hsp68 fragment was ligated into the linearised pGL4.20-DEST-*luc* via traditional ligation methods.

pGL-pZIC3-DEST-*luc*- pGL4.20-DEST-*luc* and pGl-pZIC3-*luc* were digested with *Hind*III and *Kpn*I to linearize the vector and release the *ZIC3* promoter, respectively. The *ZIC3* promoter was gel extracted and purified, and pGL4.20-DEST-*luc* purified via ammonium acetate clean up. The *ZIC3* fragment was ligated into the linearised pGL4.20-DEST-*luc* via traditional ligation methods.

pGL4.20-β-globin-NCE (wildtype) – pGL4.20-β-globin-NCE with the wildtype NCE was generated via Gateway LR cloning of the pGL-β-globin-DEST-*luc* Destination vector and pCR8-GW-TOPO-NCE. The *ZIC2 NCE* fragment was inserted into pGL-β-globin-DEST-*luc* 5' to the promoter and *luciferase* gene.

pGL4.20-Hsp68-NCE (wildtype) – pGL4.20-Hsp68-NCE with the wildtype NCE was generated via Gateway LR cloning of the pGL-Hsp68-DEST-*luc* Destination vector and pCR8-GW-TOPO-NCE. The *ZIC2 NCE* fragment was inserted into pGL-Hsp68-DEST-*luc* 5' to the promoter and *luciferase* gene.

pGL4.20-pZIC3-NCE (wildtype and mutant) – pGL4.20-pZIC3-NCE with the wildtype NCE was generated via Gateway LR cloning of the pGL-pZIC3-DEST-*luc* Destination vector and pCR8-GW-TOPO-NCE. The *ZIC2 NCE* fragment was inserted into pGL-pZIC3-DEST-*luc* 5' to the promoter and *luciferase* gene.

2.1.3 Whole Mount In Situ Hybridization (WMISH) Riboprobes

pGEM-T-Easy-*Dand5* - To make the *Dand5* (previously *Cerl2*) probe, an 888 bp fragment of exon 2 and 3'UTR from mouse *Dand5* was amplified from mouse genomic DNA (NCBI Accn. No: NR033145) using oligonucleotides RA1570 and RA1571 (Appendix Table A1.2) and TA cloned into pGEM-T-Easy using the pGEM-T-Easy Vector System (Cat. No. A1360, Promega).

pGEM-T-Easy-*Rfx3* - To make the *Rfx3* probe, a 472 bp fragment of exon 16 and the 3'UTR of mouse *Rfx3* was amplified from mouse genomic DNA (NCBI Accn. No. NM011265) using oligonucleotides RA1592 and RA1593 (Appendix Table A1.2) and TA cloned into pGEM-T-Easy using the pGEM-T-Easy Vector System.

pGEM-T-Easy-*Noto* - To make the *Noto* probe, a 386 bp fragment of the 5'UTR and exon 1 of mouse *Noto* was amplified from mouse genomic DNA (NCBI Accn. No. NM001007472.2) using oligonucleotides RA1590 and RA1591 (Appendix Table A1.2) and TA cloned into pGEM-T-Easy using the pGEM-T-Easy Vector System.

2.1.4 *PiggyBac* Transposon Constructs

pBB262-*lacZ* - p1229 and pBB262 were digested with *Xba*I and *Xho*I. The *lacZ* fragment released from p1229 was gel extracted (Section 2.1.9) and ligated into linearized pBB262. Following electrotransformation into DH5α *Escherichia coli* (*E. coli*), blue-white colony screening was used to select colonies containing the pBB262-*lacZ* construct.

pBB262-*ZIC2* NCE (wildtype and mutant) - The 608 bp wildtype *ZIC2* NCE, and each of the six mutated versions of the *ZIC2* NCE, were PCR amplified from the pCR8-GW-TOPO-NCE constructs as described in Section 2.2.13, using oligonucleotides containing *Xba*I restriction sites and homology to the pBB262 MCS at the 5' ends (RA1443 and RA1464; Appendix Table A1.2). An artificial *NotI* site was also included in oligonucleotide RA1464 to facilitate future *lacZ* cloning. pBB262 was digested with *Xba*I in the MCS to linearize the plasmid, and the NCE amplicons were cloned into pBB262 via In-Fusion Cloning (Section 2.2.12), resulting in wildtype and NCE mutant pBB262-NCE constructs (Appendix Figure A7.3).

Wildtype and mutant pBB262-NCE-*lacZ* - Following isolation and large scale prepping, pBB262-*ZIC2* NCE constructs were digested with *Not*I and *Xho*I to linearize the vector. p1229, a Bluescript II KS vector containing a 4354 bp *lacZ* reporter/SV40 PolyA cassette, was digested with *Not*I and *Xho*I to release the *lacZ*/SV40 PolyA. This was subsequently ligated into the digested pBB262-NCE constructs via traditional T4 ligation as described in Section 2.2.11. Following electrotransformation into DH5α *E. coli*, blue-white colony screening (Section 2.2.4) was used to select colonies containing wildtype and mutant pBB262-NCE-*lacZ* constructs.

pBB262-mNet – pmNetg and pBB262 were digested with *Kpn*I and *Not*I sites to release the 1617 bp mNet fragment and linearize the vector, respectively. Following gel extraction and purification, the mNet fragment was ligated into pBB262 via traditional T4 ligation as described in Section 2.2.11.

pBB262-mNet-*lacZ* – Following isolation and large scale prepping, the pBB262-mNet construct was digested with *Not*I and *Xho*I to linearize the vector. p1229, a Bluescript II KS vector containing a 4354 bp *lacZ* reporter/SV40 PolyA cassette, was digested with *Not*I and *Xho*I to release the *lacZ*/SV40 PolyA. This was subsequently ligated into the digested pBB262-mNet construct via traditional ligation. Following electrotransformation into DH5α *E. coli*, blue-white colony screening (Section 2.2.4) was used to select colonies containing the pBB262-mNet-*lacZ* construct.

2.2 Molecular Biology and Cloning

2.2.1 Electrotransformation of Plasmid DNA

Either 20 ng of plasmid DNA or 1 µL of a precipitated ligation reaction was mixed by pipetting with 25 µL of electrocompetent DH5α *E. coli* in a cold Eppendorf tube. The cells and DNA were transferred to an ice cold electrocuvette and shocked with 2.5 kV using a BioRad *E. coli* Pulser (BioRad). Immediately after shocking, 500 µL of Luria-Bertani broth (LB broth; 0.01% tryptone, 0.005% yeast extract, 0.01% NaCl; Bacto Laboratories) was added to the shocked cells and the whole volume transferred into a sterile 5 mL vented tube. The cells were allowed to recover for one hour in a 37 °C shaking incubator. 50 µL of a 1:10 dilution in LB broth was plated onto a 90 mm LB agar (LB agar; 0.010% tryptone, 0.005% yeast extract, 0.010% NaCl, 0.015% agar; Bacto Laboratories) plate containing 100 µg/mL of ampicillin, kanamycin or spectinomycin (Sigma Aldrich). The plate was incubated at 37 °C for 16 hours to allow transformed colonies to grow.

2.2.2 Heat-Shocking of Plasmid DNA

Either 10 ng of plasmid DNA or 1 µl of a precipitated ligation reaction was mixed by pipetting with 50 µL of heat-shock competent Fusion Blue cells (Clontech) in a cold Eppendorf. The cells and DNA were incubated on ice for 30 minutes, shocked at 42 °C for 60 seconds and following this, incubated on ice again for 2 minutes. 250 µL of SOC broth (Super Optimal broth with Catabolite repression; 0.02% tryptone, 0.005% yeast extract, 0.0048% $MgSO_4$, 0.0036% dextrose, 0.00005% NaCl, 0.00002% KCl; Scientifix) was added to the heat shocked cells and the whole volume transferred to a sterile 5 mL vented tube. The cells were allowed to recover for one hour in a 37 °C shaking incubator. 50 µL of a 1:10 dilution in SOC broth was plated onto a 90 mm LB agar plate containing 100 µg/mL of ampicillin, kanamycin or spectinomycin (Sigma Aldrich). The plate was incubated at 37 °C for 16 hours to allow transformed colonies to grow.

2.2.3 Bacterial Colony Screening

To analyse whether ligation reactions had produced the desired constructs, colony PCRs were performed on the resulting bacterial colonies. Colonies were picked from LB agar plates using a sterile pipette tip and were resuspended in 20 µL of sterile H_2O. A PCR mastermix was prepared

as described in Section 2.2.16, with 5 μL of the suspended bacteria used as template. Once prepared, the reaction mixture was treated as a normal PCR. The remaining bacterial suspension was stored at 4 °C until analysis was complete. Colonies positive for the required DNA were used to inoculate 5 mL of LB broth containing 100 μg/mL of either ampicillin, kanamycin or spectinomycin (Sigma Aldrich) in a vented 15 mL tube. The tubes were incubated in a shaking 37 °C incubator for 16 hours .

2.2.4 Blue-White Colony Selection

Blue-white colony selection was used to select colonies that contained a plasmid expressing the *lacZ* transgene. The surface of fresh LB agar plates containing 100 mg/mL of antibiotic were treated with 80 μL of a solution containing 0.1 M Isopropyl β-D-1-thiogalactopyranoside (IPTG; Sigma Aldrich) and 20 mg/mL X-Gal (5-bromo-4-chloro-3-indolyl β-D-galactopyranoside; Sigma Aldrich) and incubated in the dark for 30 minutes. 50 μL of electroporated or heat shocked bacteria were plated and incubated at 37 °C for 16 hours to allow colonies to grow. Cells transformed with a plasmid containing the *lacZ* transgene formed blue colonies. Following incubation, blue colonies were used to inoculate 5 mL of LB broth containing 100 μg/mL of either ampicillin or kanamycin (Sigma Aldrich) in a vented 15 mL tube. The tubes were incubated in a shaking 37 °C incubator for 16 hours.

2.2.5 Plasmid DNA isolation

Small quantities (<2 μg) of plasmid DNA was isolated from a 2-3 mL DH5α *E. coli* overnight culture with appropriate antibiotics using the High Pure Plasmid Isolation kit (Roche). Larger quantities (>2 μg) of plasmid DNA was isolated from 100-200 mL overnight culture with appropriate antibiotics using the Nucleobond Xtra Miniprep kit (Scientifix) according to the manufacturer's instructions. In both cases DNA was resuspended in ddH_2O and stored at -20 °C.

2.2.6 Agarose Gel Electrophoresis

DNA fragments were separated through horizontal gel electrophoresis. Agarose gels were created using Ultrapure Agarose (Life Technologies) and 1X TBE buffer (0.1 M Tris-HCl, 0.09 M boric acid and 0.001 M ethylenediaminetetraacetic acid [EDTA]; Amresco). DNA that ranged between 10000 bp and 500 bp was run on a 1% gel, whilst DNA between 500 bp and 100 bp was run on a 2% gel. RedSafe DNA Stain (1:20 000 dilution; ChemBio) was used to allow for visualization of the DNA. To load DNA onto the gel, 1 μL of loading dye (10X; 20% glycerol [Merck], 19.2% 0.5 M Na_2EDTA [Sigma Aldrich] and 0.001% bromophenol blue [Sigma Aldrich]) was added to every 9 μL of DNA sample, or 1 μL added to 4 μL RNA sample. DNA and RNA were run parallel to an aliquot of 1 Kb Plus DNA Ladder (100-500 ng; Life Technologies) to allow for size identification of the DNA sample and ensure the integrity of the agarose gel for both DNA

and RNA samples. Agarose gels were electrophoresed at 5-6 V/cm^{-1} until DNA and RNA bands had separated (generally 20-60 minutes). The Gel Doc XR system and the accompanying Quantity One software (Bio-Rad) were used to visualise DNA products on the gel using an ultra-violet (UV) light source, and to photograph the gel.

2.2.7 Restriction Enzyme Digests

DNA was digested with restriction enzymes (New England Biolabs [NEB]) for analysis of plasmids and as part of the cloning procedure. Digests were carried out using restriction enzymes and buffers appropriate for the DNA in question (Table 2.1). For the best results, the restriction enzymes were used at a concentration of 5 enzyme unites (U)/μg of DNA being digested (~500 ng plasmid DNA for analysis, 2-10 μg for cloning and riboprobe digestion) in 2 μL of the appropriate NEB buffer (10X), 2 μL Bovine Serum Albumin (BSA; 10X), and enough AnalaR H_2O to give a total volume of 20 μL. Digest reactions were incubated at either 37 °C or 65 °C for 1-2 hours, depending on the requirements of the enzyme. When suitable, enzyme activity was stopped by incubating the reaction at 65 °C or 80 °C for 20 minutes. Inactivated reactions were purified via gel extraction or ammonium acetate precipitation.

2.2.8 Antarctic Phosphatase Treatment

In cases where only one restriction enzyme was used to linearize plasmid DNA, Antarctic Phosphatase treatment was used to dephosphorylate the DNA and prevent unwanted religation. For the best results, the Antarctic Phosphatase (5000 U/mL; NEB) enzyme was used at a concentration of 5 U/μg of digested DNA, along with 6 μL of Antarctic Phosphatase buffer (10X; NEB). AnalaR H_2O added to make the volume up to 60 μL. Once prepared, the reaction was incubated at 37 °C for 30 minutes to ensure the DNA was completely dephosphorylated, followed by a 65 °C incubation for 5 minutes to inactivate the phosphatase. Inactivated reactions were purified via gel extraction or ammonium acetate precipitation.

2.2.9 Gel extraction and PCR clean up

Upon PCR fragment or digested plasmid DNA separation via agarose electrophoresis, the gel was placed under UV light to visualise the position of the DNA bands and the appropriate band cut out and placed into a clean 1.5 mL Eppendorf tube using a clean scalpel. The DNA was purified out of the agarose using the Nucleospin Gel and PCR Extraction kit (Scientifix) according to manufacturer's instructions.

Table 2.1: Restriction enzymes used in molecular cloning and *PiggyBac* transgenesis. All restrictions enzymes were from NEB.

Restriction Enzyme	**Buffer**	**Incubation temperature**	**Stop reaction temperature**
*Acc*I	NEB CutSmart	37°C	80°C
*Bgl*II	NEB 3.1	37°C	No
*Eco*RI	NEB CutSmart	37°C	65°C
*Eco*RV	NEB 3.1	37°C	80°C
*Hind*III	NEB CutSmart	37°C	80°C
*Kpn*I	NEB 1.1	37°C	No
*Nco*I	NEB 3.1	37°C	80°C
*Not*I	NEB 3.1	37°C	65°C
*Sac*I	NEB 1.1	37°C	65°C
*Sac*II	NEB CutSmart	37°C	65°C
*Sal*I	NEB 3.1	37°C	65°C
*Xba*I	NEB CutSmart	37°C	65°C
*Xho*I	NEB CutSmart	37°C	65°C
*Xmn*I	NEB CutSmart	37°C	65°C

2.2.10 Ammonium Acetate Precipitation

DNA that did not need to be isolated on an agarose gel was precipitated using a final concentration of 2.5 M ammonium acetate. 10X volume of cold 100% ethanol (EtOH) was then added. Following centrifuging at 18500 x *g*, 4 °C for 20 minutes to pellet the DNA, the supernatant was removed and the pellet washed with 180 µL of 70% cold EtOH for two minutes at 4 °C and 18500 x *g*. The supernatant was again removed and the pellet air-dried and resuspended in a suitable volume of AnalaR H_2O.

2.2.11 T4 DNA Ligation and Clean Up

Ligations were set up using T4 DNA Ligase (400 U/µL; Cat. No. M0202L, NEB Biolabs) and 10X T4 DNA Ligase Buffer (NEB) according to the manufacturer's instructions. Reactions were set up as either a 3:1 or 10:1 molar ratio of vector to insert, calculated with the following formula:

$$\frac{\text{Insert length (bp)}}{\text{Vector length (bp)}} \times 50 \text{ ng vector} = \text{x ng of insert required for 1:1 molar ratio}$$

The ligation reaction was either incubated at room temperature for one hour, or 16 °C overnight to increase the number of transformants.

To precipitate ligation reactions, 5 µL yeast tRNA (at 1 µg/µL), 12.5 µL of ammonium acetate (at 7.5 M) and 70 µL absolute EtOH (at -20 °C) were added to the reaction and vortexed thoroughly to mix. The reaction was immediately centrifuged at room temperature for 20 minutes, 16000 x *g*. The supernatant was discarded and replaced with 500 µL of 70% EtOH (at -20°C). The reaction was then centrifuged for 2 minutes at room temperature, 16000 x *g*, after which the supernatant was discarded, with any remaining EtOH allowed to air dry off. Once dry, the pellet was resuspended in 10 µL of AnalaR H_2O and either heat-shocked or electrotransformed into competent bacteria.

2.2.12 In-Fusion Cloning

In-Fusion cloning allows for directional cloning of DNA fragments into any vector, via a proprietary enzyme that recombines DNA fragments containing at least 15 bp of homology with the site of insertion in the linearized target vector. The In-Fusion™ Dry-Down PCR Cloning kit (Clontech) was used, according to the manufacturer's instructions, to insert the wildtype and mutant *ZIC2* NCE into pBB262 (Section 2.1) for *PiggyBac* transposon transgenic experiments and to create expression constructs for heterologous reporter assays. Briefly, 200 ng of insert and 100 ng of vector were combined with 2 µL of 5X In-Fusion HD Enzyme Premix, and the total reaction volume brought up to 10 µL with ddH_2O. The reaction was incubated at 60 °C for 15

minutes, before transferring to ice. The reaction was then transformed into Stellar™ Electrocompetent Cells or Fusion-Blue™ Competent Cells (Clontech) via heat-shock.

2.2.13 Gateway Vector Conversion

The Gateway™ Vector Conversion System with One Shot™ *ccdB* Survival Cells (Thermo Fisher Scientific; Cat. No. 11828-029) was used to convert pGL4.20 to pGL4.20-DEST (Section 2.1). Briefly, 5 μg of pGL4.20 was linearized with the restriction enzyme *EcoRV* to create blunt ends, and treated with Antarctic Phosphatase to remove 5' phosphatases and prevent vector religation. 50 ng of the dephosphorylated vector was incubated with 10 ng of Gateway Reading Frame Cassette A (Thermo Fisher Scientific) in 5X T4 DNA ligase buffer, 1U T4 DNA ligase (Thermo Fisher Scientific) and ddH_2O at a total volume of 10 μL, at room temperature for 1 hour. The reaction was then transformed into One Shot™ *ccdB* Survival Cells (Thermo Fisher Scientific) via heat-shock.

2.2.14 Gateway LR Recombination

DNA sequences were transferred from Entry vectors (pENTR) into Destination expression vectors (DEST) using Gateway LR reactions. Gateway BP and LR reactions are used by bacteriophage *Lambda phage* to insert and excise DNA sequences into/from bacterial chromosomes. For cloning purposes, the LR reaction can be triggered *in vitro* by adding bacteriophage *Lambda* recombination proteins and the *E. coli*-encoded protein Integration Host Factor. The reaction catalyses recombination between plasmids containing *att*L sites (in this case the Entry vectors) and plasmids containing *att*R sites (in this case the Destination vectors), thus the sequence between *att*L sites is transferred between the *att*R sites. The desired plasmid is selected by a combination of antibiotic resistance and the negative selection of the original Destination vector based on the presence of the lethal *ccdB* gene.

Multiple DNA sequences were transferred from pENTR1A, 2B and 3C entry vectors into various destination vectors using the Gateway LR Clonase II Enzyme mix (Invitrogen; Section 2.1), according to the manufacturer's instructions.

2.2.15 TA Cloning

The pGEM-T Easy Vector System (Promega) was used to clone multiple DNA sequences in pGEM-T-Easy, according to the manufacturer's instructions. Briefly, 50-200 ng PCR amplicon was incubated with 50 ng of pGEM-T-Easy Vector (Promega) at a 1:3 molar ratio. Additionally, 5 μL of 2X Rapid Ligation Buffer, 3U T4 DNA ligase (Promega) and ddH_2O were added at a total volume of 10 μL. Reactions were incubated at room temperature for 1 hour. Following ligation clean up, reactions were electrotransformed into competent bacteria.

2.2.16 Polymerase Chain Reaction

PCR was carried out using a Mastercycler® Vapo.protect™ or Mastercycler® Epgradient S™ (Eppendorf). PCR oligonucleotides were designed for sequencing, colony PCR and general PCR amplification (Appendix Table A1.2). To promote specificity, each oligonucleotide had a G/C clamp at the 3' end such that three of the five 3' most bases were G or C. The melting temperature (T_m) of the oligonucleotides was designed to be within 45–65 °C and oligonucleotide pairs were designed to have a T_m within 5 °C of each other.

PCR products were amplified from approximately 20-40 ng of plasmid DNA or 20 ng of PCR fragments using two oligonucleotides at a final concentration of 0.6 µM each, and a commercial PCR mastermix. Five PCR mastermixes were used, dependent on the reaction conditions (ImmoMix™ [Bioline]; BioMix Red [Bioline]; PCR ReddyMix™ Master Mix with and without KCl [Thermo Scientific], MyTaq™ HS Mix [Bioline]; Table 2.2). Betaine was used at a concentration of 1 M for DNA that was rich in dGTPs and dCTPs. A variety PCR programs were used, depending on the source of the template DNA and the annealing and melting temperatures of the oligonucleotides in each reaction. Success of the PCR was analysed via gel electrophoresis.

2.2.17 DNA sequencing

The BigDye® Direct Cycle Sequencing Kit (Life Technologies) was used to sequence DNA fragments or plasmids. Purified products were amplified using 3.5 µL 5X sequencing buffer (400 mM Tris pH 9.0, 10 mM $MgCl_2$), 1 µL BigDye Terminator v3.0 Cycle Sequencing Ready Reaction and 3.2 pm of the desired oligonucleotide (Appendix Table A1.2) in a total volume of 20 µL. Cycling conditions used were 94 ˚C for 5 minutes and 96 ˚C for 10 seconds, 50 ˚C for 5 seconds, 60 ˚C for 4 minutes for 30 cycles, 4 ˚C hold. The products were then precipitated; to each 20 µL reaction, 2 µL of 125 mM EDTA (pH 8.0), 3 µL of 3M sodium acetate (NaOAc; pH 5.2) and 50 µL of 100% EtOH were added, and samples incubated at room temperature for 15 minutes. Samples were subsequently centrifuged for 10 minutes, 3200 x *g*, 4 ^{o}C and the supernatant discarded. 250 µL of 70% EtOH was added to each sample, and then samples were again centrifuged for 10 minutes at 3200 x *g*, 4 ^{o}C. The supernatant was discarded and the samples air dried to remove any additional EtOH. Genomic sequencing was performed using the ABI 3730 sequencer (Applied Biosystems) by the Biomolecular Resource Facility (BRF; John Curtin School of Medical Research, Australian National University). The resulting sequence reads were analysed using Geneious Pro 5.5.9 (Biomatters).

Table 2.2: PCR mastermixes used for amplifying plasmid and genomic DNA.

Mastermix	Components	Brand
2X ImmoMix™	IMMOLASE™ DNA Polymerase, Stabiliser, 2 mM dNTPs, 32 mM $(NH_4)_2SO_4$, 134 mM Tris-HCl [pH 8.3], 0.02% Tween 20, 3 mM $MgCl_2$	Bioline
2X BioMix Red	BIOTAQ™ DNA Polymerase, Stabiliser, 2 mM dNTPs, 32 mM $(NH_4)_2SO_4$, 125 mM Tris-HCl [pH 8.8], 0.02% Tween 20, 3 mM $MgCl_2$, Inert dye	Bioline
2X PCR ReddyMix™ Master Mix with KCl	ThermoPrime *Taq* polymerase, 75 mM Tris-HCl (pH 8.8), 50 mM KCl, 20 mM $(NH_4)_2SO_4$, 1.5 mM $MgCl_2$, 0.01% (v/v) Tween 20, 0.2 mM dNTPs, Inert dye	Thermo Scientific
2X PCR ReddyMix™ Master Mix without KCl	ThermoPrime *Taq* polymerase, 75 mM Tris-HCl (pH 8.8), 20 mM $(NH_4)_2SO_4$, 1.5 mM $MgCl_2$, 0.01% (v/v) Tween 20, 0.2 mM dNTPs, Inert dye	Thermo Scientific
2X MyTaq™ HS Mix	MyTaq HS™ DNA polymerase, 2 mM dNTPs, MyTaq HS™ Buffer	Bioline

2.2.18 RNA Extraction and Integrity Analysis

The total RNA from embryos was extracted immediately after dissection with the RNeasy® Micro Kit (Qiagen; Cat. No. 74004) according to the manufacturer's instructions, and quantified via nanodrop. Additional DNase I treatment (Cat. No. 04716728001, 10U, Roche) at 37 °C for 20 minutes was performed to ensure no genomic DNA contamination was present. This was confirmed via standard PCR with oligonucleotides that spanned intron 1 of *Shh* (RA748 and RA749; Appendix Table A1.2), whereby no amplification indicated the samples were free from genomic DNA. RNA integrity was confirmed via agarose gel electrophoresis.

The total RNA from transfected cells was extracted 24 hours post-transfection via the NucleoSpin® RNA kit (Macherey-Nagel; Cat. No. 740955) according to the manufacturer's instructions and quantified via nanodrop. Additional DNase I treatment (Cat. No. 04716728001, 10U, Roche) at 37 °C for 20 minutes was performed to ensure no genomic DNA contamination was present. This was confirmed via standard PCR with oligonucleotides that spanned the 5'UTR of *ZIC2* (RA1595 and RA1622; Appendix Table A1.2), whereby no amplification indicated the samples were free from genomic DNA. RNA integrity was confirmed via agarose gel electrophoresis. In both cases, RNA samples were stored for up to two weeks maximum at -80 °C before RT-qPCR analysis.

2.3 Cell Culture

2.3.1 Cell Lines

Human Embryonic Kidney 293T (HEK293T) cells, a highly transfectable human epithelial-like cell line, were cultured in Dulbecco's Modified Eagle Medium (DMEM; Life Technologies) supplemented with 2 mM L-glutamine (Life Technologies), 0.1 mM non-essential amino acid solution (Life Technologies), and 10% (v/v) foetal bovine serum (FBS; Life Technologies) (referred to as supplemented DMEM herein). The cells were grown on sterile plasticware (Corning) in a humidified incubator at 37 ^{o}C and 5% CO_2.

LLC-PK1 cells, a ciliated kidney proximal tubule pig cell line, were cultured in a 1:1 ratio of Ham's Nutrient Mixture F12 (Sigma Aldrich) and DMEM (Life Technologies) that had been supplemented with 2 mM L-glutamine (Life Technologies), 0.1 mM non-essential amino acid solution (Life Technologies), and 10% (v/v) FBS (Life Technologies; 10100147) (referred to as supplemented F12:DMEM herein). The cells were grown on sterile plasticware (Corning) in a humidified incubator at 37 ^{o}C and 5% CO_2.

2.3.2 Passaging Cells

Cells grown in tissue culture plasticware (T75 cm^2 or T25 cm^2) and at 60-70% confluency were first rinsed with 1X phosphate buffered saline (PBS; Amresco) and then dissociated into a single suspension by the addition of 0.5 g/L Trypsin (Life Technologies) in 1X PBS and incubation at 37 °C for 5 minutes. Supplemented DMEM or F12:DMEM was added to inhibit the trypsin. An appropriate amount of the cell suspension, as determined by end use requirements, was transferred to new tissue culture plastic ware. Additional supplemented media was added to reach the minimum volume requirements per tissue culture flask or dish. Luciferase assays were only performed on cells between passage 11 and passage 15 to ensure the consistency of experimental repeats.

2.3.3 Transfection of Mammalian Cell Lines

Cells were transiently transfected with various plasmids using Lipofectamine™ 2000, a lipid-based transfection agent (Life Technologies). Cells to be transfected were plated into 6 well or 12 well sterile tissue culture dish (Corning; Table 2.3). If the cells were to be used for immunofluorescence or X-Gal staining, circular glass coverslips (13 mm, No. 1 thickness; ProSciTech) were added to the dish before the cells were plated. Cells were transfected when 60-80% confluent.

For each well of cells to be transfected, a total concentration of 1500-5000 ng of plasmid DNA was diluted in unsupplemented DMEM or unsupplemented F12:DMEM, mixed with Lipofectamine™ 2000 and incubated as per the manufacturer's instructions. During this time, the media on the cells was replaced with 2 mL of fresh supplemented media. The DNA-Lipofectamine™ 2000 mixture was added drop-wise to the well containing cells. The cells were returned to the 37 °C humidified incubator for a maximum of 24 hours.

2.3.4 Luciferase Assays

24 hours post transfection, the cells of each well were dissociated from the growth surface by replacing the media with fresh unsupplemented DMEM to create a cell suspension. The amount of DMEM added was calculated based on the confluency of the cells per well, with the final suspension representing a confluency of 90%. A 1:1 ratio of cells suspension to ONE-Glo Luciferase reagent (Promega, Cat. No. E6110) was created by adding equal volumes of both to a sterile 1.5 mL Eppendorf tube, per sample. The remaining cell suspension was for Western blot analysis. 100 μL of the resultant cell suspension/ONE-Glo mix was plated into three wells of a solid white tissue-culture treated 96-well plate (Corning, Cat. No. CLS3917) to create three independent internal replicates for each sample. The intensity of luminescence produce was measured by the GloMax®-96 Microplate Luminometer (Promega), or the TECAN Infinite

Table 2.3: Transfection amounts based on culture plate sizes. Separate dilutions were made for Lipofectamine™ 2000 and plasmid DNA e.g. in a 12-well plate, 4 µL of Lipofectamine™ 2000 was diluted in 100 µL of unsupplemented DMEM, whilst 1.6 µg of plasmid DNA was diluted in another 100 µL unsupplemented DMEM.

Dish type	**Cell number (100% confluency)**	**Lipofectamine™ 2000**	**Total plasmid DNA (µg)**	**Total well volume**
12 well plate	3.8×10^5	4 µL/well	1.6	1.0 mL
6 well plate	9.5×10^5	10 µL/well	4.0	2.5 mL

M1000 Pro. The mean relative luciferase activity and standard deviation were calculated from three internal repeats (using Microsoft Excel). Where pooled data is shown, data represents at least three independent repeats, with luminescence values from one representative experiment shown. To avoid the potential problem of position bias by the luminometer, sample order on the plate was randomized for each experimental repeat.

2.3.5 Serum Starvation

To increase the number of cilia in LLC-PK1 cells, the cells were serum starved. 24 hours post transfection, supplemented F12:DMEM was aspirated from cells and replaced with unsupplemented F12:DMEM. 48 hours post transfection, media was aspirated and cells were taken through the immunofluorescence protocol as outlined in Section 2.5.2.

2.4 Reverse Transcription Quantitative PCR (RT-qPCR)

RT-qPCR oligonucleotide sequences were selected over intron-exon boundaries whenever possible to limit genomic DNA amplification (Appendix Table 1.2, Table 2.4). One-step RT-qPCR was performed on 50 ng of RNA per sample per well of a 96 well sterile PCR plate, using the SensiFAST SYBR Hi-ROX One-Step Kit (Bioline) at the recommended conditions (Table 2.5). Plates were sealed with UltraClear Sealing Film (Fisher Biotec) and centrifuged at 180 x *g* for 1 minutes. cDNA conversion and melt curve analysis were performed using a StepOnePlus™ Real-Time PCR System (Applied Biosystems®) and StepOne software (version 2.2.2; Applied Biosystems®). For quantification, the assay for each gene consisted of three internal replicates per gene per sample. At least three independent experiments were performed for each RT-qPCR. Mean values were calculated and the v5.5.9. Methylation and acetylation values were analysed in Microsoft Excel to correlate enrichment with the *ZIC2* 3'UTR sequence.

2.4.1 Statistics

Luciferase assays: The standard deviation (S. D.) of the internal repeats of the representative experiment was calculated using Microsoft Excel. Pooled data was analysed using Genstat software to perform either an unpaired Student's T-Test or two-way Analysis of Variance (ANOVA) with Fischer's unprotected post hoc test on a minimum of three external repeats, and to calculate the standard error of the mean (S.E.M.).

(C_T) value used for analysis.

2.4.2 Oligonucleotide Efficiencies

All of the RT-qPCR oligonucleotide pairs were tested for their amplification efficiency prior to sample analysis. PCR efficiencies were calculated using the slope of a calibration curve as

Table 2.4: RT-qPCR oligonucleotide efficiencies. Oligonucleotide sequences can be seen in Appendix Table A1.2.

Gene		Oligonucleotide	Species	Product size (bp)	Slope	R^2	Efficiency
H2afz	F	RA1719	Mouse	202	-3.30	0.99	101.00%
	R	RA1720					
mP0	F	RA1780	Mouse	109	-3.35	0.99	99.00%
	R	RA1781					
mP1	F	RA99	Mouse	101	-3.32	0.99	100.26%
	R	RA1782					
mP2/ZIC2 3'UTR	F	RA1714	Mouse/ Human	101/105	-3.23	0.99	103.88%
	R	RA1715					
mP3	F	RA1716	Mouse	109	-3.30	0.99	100.85%
	R	RA1717					
Ubc	F	RA1730	Mouse	112	-3.33	0.99	99.47%
	R	RA1731					
Zic2	F	RA252	Mouse	178	-3.27	0.98	102.34%
	R	RA96					
ABCE1	F	RA1606	Human	234	-3.25	0.98	103.06%
	R	RA1607					
FOXA2	F	RA1766	Human	90	-3.16	0.98	107.25%
	R	RA1767					
FOXJ1	F	RA1768	Human	93	-3.24	0.99	103.63%
	R	RA1769					
PPP1R8	F	RA1604	Human	123	-3.33	0.99	99.70%
	R	RA1605					
ZIC2	F	RA1177	Human	154	-3.43	0.99	95.61%
	R	RA1247					

Table 2.5: One-step RT-qPCR cycling and melt curve conditions. Conditions are optimised for the SensiFAST SYBR Hi-ROX One-Step Kit (Bioline).

Program step	**Temperature**	
Initial hold	45 °C 20 min	
Initial denaturation	95 °C 2 min	
Denature	95 °C 5 sec	X 40
Anneal	56 °C 10 sec	
Extend	72 °C 10 sec	
Melt curve	95 °C 15 sec	
	60 °C 1 min, +0.3 °C per minute until 95 °C is reached	
Polish	95 °C 15 sec	
Stop	4 °C	Hold

described in Bustin et al (2009). Efficiency was assessed using the standard curve method. RNA samples with a starting concentration of 100 ng/µL were serially diluted 10-fold. These serial dilutions (100 ng/µL, 10 ng/µL, 1 ng/µL, 0.1 ng/µL, 0.01 ng/µL and 0.001 ng/µL) were then run through the RT-qPCR protocol in triplicate. For each target sequence, the mean C_T obtained during the amplification of each dilution was plotted against the log of the input concentration of RNA to create a standard curve. The linear regression and coefficient of determination (R^2) of the standard curve were calculated using the Microsoft Excel LINEST function. The efficiency of the assay is calculated from the slope of the regression line of the standard curve, using the formula: $E = 10^{-1/slope}$.

An oligonucleotide pair was considered efficient if they produced an R^2 value ≥0.98 and, ideally, a slope value of -3.32 resulting in an efficiency of 100% (where 100% represents doubling of the PCR product every cycle). Due to normal experimental variability, efficiency values between 90-110% were also acceptable for the conditions of this RT-qPCR assay. RT-qPCR oligonucleotides and their efficiencies are shown in Table 2.4.

2.4.3 $2^{-\Delta\Delta CT}$ Calculations

The StepOnePlus™ Real-Time PCR system automatically records the PCR cycle number where the fluorescence signal of a sample significantly rises above the background fluorescence. This C_T is inversely proportional to the amount of input target nucleic acid and thus allows the starting amount of the target sequence to be quantified. The StepOnePlus™ RT-PCR system plots the negative first derivative of the dissociation curve for each sample, producing a dissociation peak graph from which the T_m: (the point at which 50% of the DNA is single stranded) of the DNA in the sample can be identified. A single peak observed on the dissociation curve indicates there is a single DNA product in the reaction, while multiple peaks indicate multiple PCR products are present (e.g. from genomic contamination or oligonucleotide dimers). The C_T values from samples with melt curves showing multiple points were excluded from further analysis.

The $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001) was used to calculate the fold increase in the expression of the target sequence. The mean and standard deviation of the C_T values for the technical replicates of each sample was calculated with Excel. Replicates with a C_T value that varied from the mean C_T value by more than 1 cycle were discarded as outliers and a new mean C_T calculated from the remaining two replicates. The normalization process was only carried out between samples analysed in the same RT-qPCR run using the following equations:

$$\Delta CT = CT \text{ (target gene)} - CT \text{ (reference gene)}$$

$$\Delta\Delta CT = \Delta CT \text{ (test sample)} - \Delta CT \text{ (control sample)}$$

$$\text{Fold change} = 2^{-\Delta\Delta CT}$$

The mean C_T of the target of interest was normalised to that of a reference gene (*H2afz* and *Ubc* for mouse RNA, *PPP1R8* and *ABCE1* for human RNA). The normalised CT (ΔCT) of samples from control samples is subtracted from that of test samples, to account for endogenous expression (ΔΔCT). The fold change in gene expression is then calculated simply as $2^{-\Delta\Delta CT}$.

2.5 Cell Staining

2.5.1 X-Gal Staining of Cultured Cells

For X-Gal staining, cells were grown onto glass coverslips as described in Section 2.2.3. 24 hours post transfection, cells were fixed in 4% paraformaldehyde (PFA) in 1X PBS for 30-60 min. The cells were rinsed in 1X PBS before being incubated overnight in X-Gal stain (5 mg/mL 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside in dimethylformamide, Sigma Aldrich; 1 mM $MgCl_2$, BDH; 5 mM potassium ferrocyanide, Sigma Aldrich; 5 mM potassium ferricyanide, Sigma Aldrich; 1X PBS) at 37 °C.

Following incubation in stain, cells were rinsed with 1X PBS. The coverslips were mounted onto glass slides (76 x 26 mm; ProSciTech) with anti-fade mounting agent (2.4 mM n-propyl gallate, NPG [Sigma Aldrich] in a 1:1 solution of glycerol and 1X PBS). The edges of the coverslips were sealed with nail polish to prevent the cells from drying out.

2.5.2 Immunofluorescence

For immunofluorescence staining, cells were grown onto glass coverslips (no.1 thickness, ProSciTech). 24 hours post transfection, cells were fixed in 4% PFA (ProSciTech) in 1X PBS for 30-60 min. The cells were rinsed in 1X PBS three times before being permeabilised in 0.25% Triton X-100 (Sigma Aldrich) in 1X PBS. The cells were again rinsed three times in 1X PBS before being blocked for one hour at room temperature in a skim milk blocking solution (5% skim milk powder, Diploma; 1 X PBS) at 4 °C.

To detect proteins of interest, primary antibodies were diluted in skim milk blocking buffer (Table 2.6) for 1 hour at room temperature. The diluted primary antibody solution was applied to the fixed and blocked cells for one hour in a humidified chamber. Subsequently, the cells were rinsed with blocking solution to remove unbound antibody. Fluorophore conjugated secondary antibodies (used to detect the primary antibodies; Table 2.7) were diluted in blocking buffer and applied to the primary antibody stained cells for one hour in a darkened humidified

Table 2.6: Primary antibodies used in immunostaining.

Target Protein	Company; Cat. No.	Species	Dilution used
V5	Sapphire Biosciences; R960-25	Rabbit, polyclonal	1:500
Acetylated Tubulin	Sigma Aldrich; T7451	Mouse, monoclonal	1:200
C-Myc	Thermo Fisher Scientific; 132500	Rabbit, monoclonal	1:500

Table 2.7: Secondary antibodies used in immunostaining.

Secondary Antibody	Species	Target Species	Company	Dilution	Absorption
Alexa^{488}	Donkey	Rabbit	Thermo Fisher Scientific; A21206	1:500	495 nm
Alexa^{594}	Donkey	Mouse	Thermo Fisher Scientific; A21203	1:500	590 nm

chamber. The cells were rinsed with 1X PBS six times to remove unbound secondary antibody and the coverslips mounted onto glass slides with anti-fade mounting agent (2.4 mM NPG [Sigma Aldrich] in a 1:1 solution of glycerol and PBS). The edges of the coverslips were sealed with nail polish to prevent the cells from drying out.

2.6 Cell Lysis and Western Blotting

2.6.1 Nuclear/Cytoplasmic Protein Extraction

Protein for western blotting was extracted from transfected cells and fractionated using the NE-PER Nuclear and Cytoplasmic Reagents (Pierce) according to the manufacturer's instructions, but altering the volumes. Media was removed from cells in a 12 well plate and a cell suspension made in 500 µL fresh, unsupplemented DMEM. Cells were pelleted by centrifugation at 18000 x *g* and the media removed and replaced with 50 µL of ice cold CER I buffer, with Protease Inhibitor (cOmplete, EDTA-free Protease Inhibitor, Roche) and Phosphatase Inhibitor (PhosStop Phosphatase Inhibitor Cocktail, Roche) added at final concentrations of 1X. Cells were vortexed thoroughly and incubated for 10 minutes on ice. Following incubation, 2.25 µL of CER II was added and cells were again vortexed followed by incubation on ice for one minute and a second vortex. Cells were then pelleted by centrifugation at 18000 x *g* for 15 minutes at 4 ^{o}C, and the resulting cytoplasmic supernatant transferred to a new 1.5 mL Eppendorf tube. 25 µL of NER buffer, again containing 1X Protease Inhibitor and Phosphatase Inhibitor, was added to the cell pellet. The cells were resuspended by vortexing thoroughly and the samples incubated on ice for 40 minutes, vortexing every 10 minutes to mix. The samples were pelleted by centrifugation at 18000 x *g*, for 25 minutes at 4 ^{o}C and the resulting nuclear supernatant transferred to a new 1.5 mL tube.

To prepare the cytoplasmic and nuclear fractions for Sodium Dodecyl Sulfate-polyacrylamide gel electrophoresis (SDS-PAGE), DTT (dithiothreitol; Sigma-Aldrich, at 20X) and NuPAGE LDS Sample Buffer (Invitrogen; 4X) were added to the cell lysates to give a final concentration of 50 mM and 1X respectively. The samples were heated at 90 ^{o}C for 5 minutes to denature the proteins. Denatured lysates were then either used immediately or stored at -80 ^{o}C.

2.6.2 SDS-PAGE

Protein from cytoplasmic and nuclear cell lysates was separated by molecular weight (MW) on an acrylamide gel. Polyacrylamide 10% running gel mixes (Table 2.8) were made and cast in the Mini PROTEAN® casting apparatus (Bio-Rad) while still liquid. Running gels were overlaid with hydrated butanol (Sigma-Aldrich) and left at room temperature to polymerise for 1 hour. Following polymerisation, the butanol was removed, the running gels were overlaid with a

Table 2.8: Reagents and volumes for a 10% running polyacrylamide gel. Volumes are for one gel of 1.5 mm thickness.

Reagent	Volume
40% Acryl:Bisacryl (37.5:1) (Sigma-Aldrich; Cat. No. A7168)	3.13 mL
1.5 M Tris pH 8.8 (Sigma-Aldrich; Cat No. 252859)	3.13 mL
10% SDS (Sigma-Aldrich; Cat. No. L4390)	125.00 μL
ddH_20	6.00 mL
10% APS (Sigma-Aldrich; Cat. No. A3678)	125.00 μL
TEMED (Sigma-Aldrich; Cat. No. T9281)	12.50 μL

Table 2.9: Reagents and volumes for a 3.75% stacking polyacrylamide gel. Volumes are for one gel of 1.5 mm thickness.

Reagent	Volume
40% Acryl:Bisacryl (37.5:1) (Sigma-Aldrich; Cat. No. A7168)	235.00 μL
0.5 M Tris pH 8.8 (Sigma-Aldrich; Cat No. 252859)	625.00 μL
10% SDS (Sigma-Aldrich; Cat. No. L4390)	25.00 μL
ddH_20	1.59 mL
APS (Sigma-Aldrich; Cat. No. A3678)	50.00 μL
TEMED (Sigma-Aldrich; Cat. No. T9281)	5.00 μL

3.75% stacking gel (Table 2.9) and a comb was inserted to create wells for loading. The gels were again left to polymerise for 1 hour.

Once polymerisation was complete, the gels were placed into a BIORAD Mini PROTEAN® Tetra Cell apparatus filled with 1X SDS-PAGE Running Buffer (192 mM Glycine, 24.9 mM Tris base and 3.47 mM SDS). 5-30 µL of each sample were loaded into the separate wells of the gel. Alongside these samples, 8 µL of PageRuler Prestained Protein Ladder (Fermentas) was added to visualise the protein separation during gel running and to determine approximate size of proteins after western blotting. Gels were run at 100 V for 1.5-2.5 hours to separate the proteins, using the separation of the Protein Ladder as a guide for running time.

2.6.3 Wet Transfer

Once the gel was run to an appropriate extent, the protein was transferred onto a membrane using wet transfer. The stacking portion of the gel was discarded and the running gel placed on a piece of polyvinylidene difluoride (PVDF) membrane, which had been activated in 100% methanol (MeOH). The gel and membrane were sandwiched between four pieces of blotting paper (3 mm chromatography paper, Whatman®) and two sponges within a gel holder cassette, with all components submerged in 1X Towbin's Buffer (190 mM Glycine, 24 mM Tris base and 20% MeOH) to allow the gel to equilibrate and to reduce air bubbles. The gel holder cassette (containing the gel, PVDF, blotting paper and sponges) was transferred to a mini Trans-Blot Module (Bio-Rad) within a gel tank filled with 1X Towbin's Buffer, and run at 15 V overnight (~16 hours) to transfer the protein to the PVDF membrane. During this time, the tank was surrounded by ice to minimise heating.

2.6.4 Western Blotting

Once transfer was complete the, PDVF membrane (to which the protein is now bound) was removed from the transfer apparatus and blocked in blocking buffer (5% skim milk [Diploma], 0.02% Tween 20 in PBS) for at least 1 hour with agitation. Following blocking, the membrane was exposed to a primary antibody (Table 2.10) diluted in 3 mL of blocking buffer for 2-2.5 hours. The membrane was then washed with blocking buffer six times, with each wash lasting 5-10 minutes. Following washing, the membrane was exposed to the appropriate Horse Radish Peroxidase (HRP) conjugated secondary antibody (Table 2.11) diluted in 3 mL of the blocking buffer for 1-1.5 hours and then washed with 0.02% Tween in 1X PBS six times, with each wash lasting 5-10 minutes.

The membrane was incubated in SuperSignal West Pico Chemiluminescent Substrate (Pierce) for 5 minutes in darkness. The SuperSignal West Pico Chemiluminescent Substrate contains a

Table 2.10: Primary antibodies used in western blotting.

Target Protein	Company; Cat. No.	Species	Dilution used
GAPDH	Abcam; ab8245	Mouse, monoclonal	1/1000
HA-probe (Y-11)	Santa Cruz; sc-805	Rabbit, polyclonal	1/200
Tata binding protein (TBP)	Abcam; ab818	Mouse, monoclonal	1/2000
V5	Invitrogen; R960-25	Mouse, monoclonal	1/3000

Table 2.11: Secondary antibodies used in western blotting.

Species	Target Species	Company; Cat. No.	Dilution used
Rabbit	Mouse	Invitrogen; 616520	1/5000
Donkey	Rabbit	Invitrogen; A21206	1/500

substrate upon which the HRP tag of the secondary antibody can act to produce a chemiluminescent signal. Following incubation, the membrane was immediately exposed to a charge-coupled camera via the ImageQuant™ LAS4000 biomolecular imager (GE Healthcare Life Sciences) for 10 seconds - 10 minutes (dependent on the experiment and antibody used) and a digital image produced.

2.7 Mouse Husbandry, Strains and Alleles

Wildtype embryos were taken from the C3H/HeH colony, which was maintained by continuous backcross to C3H/HeH. The *Kumba* (*Ku*) allele of *Zic2* (Brown *et al.*, 2000; Elms *et al.*, 2003; Warr *et al.*, 2008) was maintained on two distinct backgrounds by continuous backcross to either C3H/HeH or 129/SvEv mice. In both cases, mice from backcross 10 or beyond were used for analysis. Gene expression and phenotype were found to be identical between the two backgrounds for both $Zic2^{Ku/+}$ and $Zic2^{Ku/Ku}$ embryos at embryonic stages 7.0 – 9.5 dpc and analysis of embryos was performed using mice derived from both colonies. Mice were maintained in a light cycle of 12 h light: 12 h dark, the midpoint of the dark cycle being 12 A.M.

2.8 Embryo Collection

Embryos were collected from pregnant dams from 6.5 dpc onwards. Noon (12 P.M.) on the day of the appearance of the vaginal plug was designated 0.5 dpc. Embryos were dissected from the uterus and decidua in a solution of 10% FBS (Life Technologies) in 1X PBS. All membranes surrounding the embryos were dissected away; the yolk sac and amnion from mutant litters were taken for genotyping. Embryos were staged according to Downs and Davies (Downs and Davies 1993).

For WMISH, embryos were fixed overnight in 4% PFA in 1X PBS at 4 °C, and then dehydrated through a MeOH:PBS series (25%, 50%, 70% and 100% MeOH, respectively). Embryos were stored in 100% MeOH at -20 °C until WMISH was performed. For RT-qPCR, embryos were pooled based on their age. The total RNA was extracted immediately as per Section 2.2.18.

2.9 Genotyping

2.9.1 Genomic DNA Lysis

Adult mice were genotyped using ear biopsy DNA. A ~20 mm^2 notch of ear was taken from each mouse. To each ear notch sample, 50 μL of lysis solution (50 mM Tris-HCl pH 8.5, 1 mM EDTA, 5.0% Tween 20) and 2 μL of Proteinase K (PK; 10 mg/mL in dH_2O) was added and incubated at 55 °C for 60 minutes, followed by 95 °C for 10 minutes to inactivate the PK. Tissue debris was pelleted by centrifuging at 2000 x *g* for 5 minutes, and each sample diluted 1:10 in AnalaR H_2O.

Embryos were genotyped using a fragment of extra ectoplacental cone (7.5 dpc), amnion or yolk sac (≥ 8.5 dpc). To each embryo sample, 10 µL of lysis solution and 2 µL of PK was added and samples incubated at 55 °C for 25 minutes and 95 °C for 5 minutes to inactivate the PK. Tissue debris was pelleted by centrifuging at 2000 x *g* for 5 minutes, and each sample diluted 1:4 in AnalaR H_2O.

2.9.2 Genotyping PCRs

To genotype transient *PiggyBac* transgenic embryos, oligonucleotide RA1058 was paired with oligonucleotide RA1547 (pBB262-NCE-*lacZ* wildtype and mutant), RA1548 (pBB262-*lacZ*) and RA1059 (pBB262-mNet-*lacZ*). All reactions were carried out in a 10 µL volume with a final concentration of 0.6 µM of each oligonucleotide and 2X PCR ReddyMix™ Master Mix without KCl (Table 2.2). 1 M Betaine was included for pBB262-NCE-*lacZ* genotyping. PCRs to be analysed via gel electrophoresis were performed in 96-Well Clear, Flat Top PCR plates (Axygen; Cat. No. PCR-96-FLT-C). All PCRs were performed in an Eppendorf Mastercycler® using the TD70 Touchdown PCR programs (Table 2.12) and the products run on a 1.5% agarose gel. The production of a 443 bp (pBB262-mNet-*lacZ),* 561 bp (pBB262-NCE-*lacZ* wildtype and mutant) and 366 bp (pBB262-*lacZ*) were deemed a positive genotype.

Genotyping of CRISPR 3'UTR mutant embryos was performed by Nay Chi of the JCSMR Transgenesis Facility. Briefly, all reactions were carried out in a 50 µL volume with a final concentration of 0.6 µM of each oligonucleotide (RA1778 and RA1779, Appendix Table A1.2), 2X MyTaq™ HS Mix (Bioline; Table 2.2) and 1 M Betaine. PCRs to be analysed via gel electrophoresis were performed in 96-Well Clear, Flat Top PCR plates (Axygen; Cat. No. PCR-96-FLT-C). All PCRs were performed in an Eppendorf Mastercycler® using a hot start PCR program according to the manufacturer's instructions. The products run on a 1% agarose gel; an amplicon of 1019 bp indicated a wildtype embryo.

2.9.3 High Resolution Melt Assay (HRMA)

Genomic DNA extracted from ear notches of adult mice and embryo tissue were genotyped using HRMA, as described by (Thomsen *et al.*, 2012). PCRs were carried out using ImmoMix (Bioline; Cat. No. BIO-25020) and included the LC Green® Plus+ Melting Dye (Idaho Technology Inc.; Cat. No. BCHM-ASY-0005). All reactions were carried out in a final volume of 10 µL with ~30 ng of digested ear notch DNA, with a final concentration of 0.6 µM of each oligonucleotide (Appendix Table A1.2). Reactions were set up in Hard-Shell® 96-well PCR Plates (BioRad; Cat. No. HSP-9665) covered with Axygen Microplate Sealing Film (Fisher Scientific; Cat. No. UC500). To avoid evaporation during the HRMA process, each reaction was covered with ~10 µL of mineral oil prior to PCR. All PCRs were performed in an Eppendorf Mastercycler® using three Touchdown

PCR programs: TD60, TD65 and TD70 (Table 2.12). On completion of a PCR reaction, the 96-well plate containing PCR products was placed directly into a Light Scanner HR 96 (Idaho Technologies Inc.) and samples melted from 60 °C to 95 °C at a rate of 0.1 $secs^{-1}$. The LC Green® dye specifically binds to double-stranded DNA and emits fluorescence that is captured by the Light Scanner HR 96 instrument. As temperature increases double-stranded DNA is converted to single-stranded, which dissociates LC Green from DNA, resulting in a decrease in fluorescence. Since melting of DNA is dependent on sequence and length, each amplicon has a unique melt profile. The data were analysed with LightScanner software (Idaho Technologies Inc.).

Kumba wildtype and heterozygotes adults were genotyped with oligonucleotides RA247 and R248 (Appendix Table A1.2), whilst mNet wildtype and heterozygote adults and embryos were genotyped with RA1058 and RA1059. A *Shh* control was also amplified (RA748 and RA749) to ensure the integrity of mNet embryo gDNA.

2.9.4 Allelic Discrimination Assay (ADA)

ADA was used to genotype *Kumba* embryo DNA, as $Zic2^{Ku/Ku}$ homozygotes cannot be distinguished via HRMA. A TaqMan® Universal PCR Master Mix (Life Technologies; Cat. No. 4304437), along with TaqMan® MGB probes (Applied Biosystems) corresponding to the wildtype and *Kumba* (*Ku)* alleles of *Zic2* were used to visualize the genotype of each embryo. The sequence of the probes are given in Table 2.13. The wildtype probes were modified at the 5' end with a 6-FAM (6-carboxyfluorescein) tag and the mutant probes with VIC®. The 3' end of each probe was modified with a non-fluorescent quencher and a minor groove binding moiety. The probes were used at half the recommended concentration (200 nM). All reactions were carried out in a final volume of 10 µL with ~30 ng of digested embryo DNA in the presence of 0.9 µM of oligonucleotides RA247 and RA248 (Appendix Table A1.2). Reactions were set up in 96-well Half-Skirted PCR Microplates (Axygen®; Cat. No. PCR-96-LP-AB-C) covered with Axygen Microplate Sealing Film (Fisher Scientific; Cat. No. UC500) and performed using the StepOnePlus™ Real-Time PCR System (Applied Biosystems®). The StepOne Software (version 2.2.2; Applied Biosystems®) was used to run the assay using the following conditions: an initial pre-PCR read at 60 °C for 30 seconds to record background fluorescence, followed by 95 °C for 10 minutes to denature the template and a cycling stage of 95 °C for 15 seconds and of 60 °C for 1 minute for 50 cycles. A post-PCR read was performed at 60° C for 30 seconds to collect data after completion of the PCR. Data was analysed using the same software that records the pre- and post-PCR reads and calculates normalized dye fluorescence (ΔRn) from the wildtype and mutant alleles as a function of cycle number. Based on this data the software called the sample as homozygous for either wildtype or mutant allele, or heterozygous with both alleles.

Table 2.12: Touchdown PCR cycling conditions.

Program:	**TD60**		**TD65**		**TD70**	
Initial denaturation	94 °C 4 minutes		94 °C 4 minutes		94 °C 4 minutes	
Denature	94 °C 30 seconds	X 29	94 °C 30 seconds	X 19	92 °C 30 seconds	X 20
Anneal	60 °C, decreasing by 0.5 °C per cycle, 30 seconds		65 °C, decreasing by 0.5 °C per cycle, 30 seconds		70 °C, decreasing by 0.5 °C per cycle, 30 seconds	
Extend	72 °C 30 seconds		72 °C 30 seconds		72 °C 1 minute	
Denature	94 °C 30 seconds	X 19	94 °C 30 seconds	X 29	92 °C 30 seconds	X 20
Anneal	45 °C 30 seconds		55 °C 30 seconds		60 °C 30 seconds	
Extend	72 °C 30 seconds		72 °C 30 seconds		72 °C 1 minute	
Polish	72 °C 7 minutes		72 °C 7 minutes		72 °C 7 minutes	
Stop	4 °C	Hold	4 °C	Hold	4 °C	Hold

Table 2.13: ADA probes for genotyping of *Kumba* mice.

Probe name	**5' Tag**	**Probe sequence**	**3' Quencher**
Zic2 WT	6-FAM	CGA GGG CTG TGA CC	MGBNFQ
Zic2 Ku	VIC	TTC GAG GGC AGT GAC	MGBNFQ

2.10 Embryo Hybridization and Staining

2.10.1 WMISH Riboprobe Synthesis from Plasmid DNA

Riboprobes were synthesized from plasmid DNA as shown in Table 2.14. *Dand5, Rfx3* and *Noto* were cloned into pGEM-T-Easy before being synthesised (Section 2.1.3). All other riboprobes made from plasmid DNA were already in plasmid form in the Arkell laboratory. Plasmid DNA was linearized using restriction enzymes 5' to the probe, shown in Table 2.14.

Antisense RNA probes were *in vitro* transcribed from 1 µg of plasmid DNA. T7, T3 or SP6 RNA Polymerase (Roche) and Digoxygenin Labelling Mix (Roche) were used according to the manufacturer's instructions and with the addition of RNase Inhibitor (Roche). After synthesis, the DNA template was digested using 20 U of DNase I (RNase free, Roche). All reactions were stopped by the addition of EDTA at a final concentration of 16 mM. The RNA was precipitated overnight at -20°C by the addition of LiCl to a final concentration of 0.36 M and 2.5X volumes of 100% EtOH. The RNA was pelleted by centrifugation at 20800 x *g* for 20 minutes at 4 °C, and the supernatant removed. The pellet was washed with 200 µL of 70% EtOH and air dried. It was then resuspended in 100 µL of AnalaR H_2O. A 5 µL aliquot was electrophoresed on a 1% agarose/1X TBE gel to check the yield and degradation of the RNA probe. Probes were stored at -20 °C until use.

2.10.2 Whole Mount In Situ Hybridization

WMISH was performed as previously described in Wilkinson (1992) using the hybridization conditions of Rosen and Beddington (1993) (Rosen and Beddington, 1993; Wilkinson, 1992). For 12.5 dpc embryos, wash times were extended to 15 minutes and WMISH was performed on a benchtop nutating mixer to decrease trapping. PK and 4% PFA steps were also doubled. All WMISH embryos were destained in 1X PBT (1X PBS; Tween 20) and post-fixed in 4% PFA in PBS for one hour at room temperature before being photographed as described in Section 2.13.1.

2.10.3 X-Gal staining of whole embryos

Dissected embryos were fixed in 4% PFA in 1X PBS for 20 minutes at room temperature. Following this, embryos were rinsed in 1X PBS and equilibrated in a sodium phosphate wash buffer (2 mM magnesium chloride, 0.01% deoxycholate, 0.02% NP40, 98 mM sodium phosphate buffer [23 mM monobasic sodium phosphate, 77 mM dibasic sodium phosphate, pH 7.3, Sigma]). X-Gal stain (10 mg/mL 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside in dimethylformamide, Sigma Aldrich; 40 mM potassium ferrocyanide, Sigma Aldrich; 40 mM

Table 2.14: Riboprobes used for WMISH. Probes correspond to constructs listed in Appendix Table A1.1. Restriction enzymes used to linearise plasmid DNA, and RNA polymerases are indicated.

Probe	Restriction Enzyme	Transcription Polymerase	Reference
Cdx2	*Eco*RV	T7	Stefan Broer, ANU
Dand5	*Sac*II	SP6	(Barratt *et al.*, 2014)
Dkk1	*Sal*I	T7	(Ip *et al.*, 2014)
Foxa2	*Hind*III	T7	(Sasaki and Hogan, 1993)
Foxj1	*Sal*I	T3	(Cruz *et al.*, 2010)
Lefty1/2	*Xho*I	T7	(Meno *et al.*, 1996)
Lhx1	*Hind*III	T7	(Shawlot and Behringer, 1995)
Nodal	*Eco*RI	T3	(Conlon *et al.*, 1994)
Noto	*Sal*I	T7	(Barratt *et al.*, 2014)
Nppa	*Eco*RI	T3	Dominic Norris, MRC Harwell, UK
Pitx2	*Sac*I	T3	(Ryan *et al.*, 1998)
Pkd1l1	*Eco*RI	T7	(Field *et al.*, 2011)
Rfx3	*Nco*I	SP6	(Barratt *et al.*, 2014)
Shh	*Xba*I	T7	(Echelard *et al.*, 1993)
Sox2	*Acc*I	T3	(Wood and Episkopou, 1999)
Sox3	*Not*I	T7	(Wood and Episkopou, 1999)
Sox17	*Not*I	T7	Dominic Norris, MRC Harwell, UK
Zic2	*Sal*I	T7	(Elms *et al.*, 2003)
Zic3	*Nco*I	SP6	(Elms *et al.*, 2004)
Zic5	*Sal*I	T7	Arkell laboratory

potassium ferricyanide, Sigma Aldrich; 200 mM Tris-HCl pH 7.4) was diluted in the sodium phosphate wash buffer and embryos were incubated in stain overnight in a dark 37 °C shaking incubator. Embryos were destained in 1X PBT (1X PBS; Tween 20) and post-fixed in 4% PFA in PBS for one hour at room temperature before being photographed as described in Section 2.13.1.

2.11 *PiggyBac* Transposon Transgenesis

2.11.1 Generation of *PiggyBac* Transposase mRNA

To transcribe the *PiggyBac* transposase for zygote microinjections, pBB232 was linearized with *Xba*I (Table 2.1). The transposase mRNA was transcribed with T3 RNA Polymerase and the mMessage mMachine T3 Kit (Ambion) no more than 24 hours before zygote microinjections. The resulting mRNA was purified with the MEGAclear Transcription Clean-Up kit (Ambion), and mRNA quality and yield assessed via agarose gel electrophoresis and nanodrop. Only transposase mRNA with clear bands, high yield and no gDNA contamination was used in the subsequent microinjections.

2.11.2 *PiggyBac* Pronuclei Microinjection and Implantation

Embryo donor FVB/NJ mice were first injected with Pregnant Mare Serum Gonadotrophin, followed by Human Chorionic Gonadotrophin 48 hours later to induce super ovulation. They were then immediately placed with fertile stud males and fertilised zygotes were harvested 12 hours post-conception. *PiggyBac* transposon mRNA and pBB262 transgene constructs (pBB262-mNet-*lacZ*, pBB262-*lacZ* and wildtype pBB262-NCE-*lacZ*) linearized with *XmnI* (Table 2.1) were co-injected into the pronucleus by Gene Elliot at the NIH Transgenic Core Facility. Following injection, zygotes were incubated overnight (37 °C, 5% CO_2) until they reached the 1-2 cell stage. Unviable embryos were discarded, whilst viable embryos were implanted into the right uterine infundibulum of anesthetised pseudopregnant FVB/NJ mice and left to develop until the desired dissection time point. Approximately 15 1-2 cell stage embryos were implanted in each female.

2.12 CRISPR-Cas9 Mutagenesis

2.12.1 Guide Design

Zic2 3'UTR deletion CRISPR guide design was primarily performed by the JCSMR Transgenesis Facility (ANU). Guide strands were designed using the Zhang CRISPR Design Tool (Zhang Lab and MIT, 2015). Sequences of 150 bp 5' and 3' to the *Zic2* 3'UTR enhancer region were analysed for appropriate candidate sites. Appropriate guide strands were identified based on the presence of a 5'-NGG PAM sequence immediately downstream (3') of the candidate site. Candidate guide strands were deemed ideal if: the candidate was 20 bp in length, >35% and <80% of the

nucleotides were G or C, candidates had a small number of off target sites, and the off-target sites that were present had three or more mismatches to the guide strand. Due to the location of the guide within the *Zic2* 3'UTR, however, these criteria were not always met. A G was added to the 5' of those guide sequences that did not already have a G as the first base pair.

2.12.2 Preparation of CRISPR Guide Plasmid DNA and mRNA

Guide oligonucleotides that were purchased as DNA were cloned into pTOPO-Blunt-II by Nay Chi of the JCSMR Transgenesis Facility. Upon largescale isolation, plasmid DNA was injected directly into fertilized zygotes, as described in Chapter 7. Guides that were purchased as mRNA were directly injected into fertilized zygotes.

2.12.3 CRISPR Guide mRNA and Cas9 Microinjection and Implantation

Embryo donor C57BL/6-NCrl mice were first injected with Pregnant Mare Serum Gonadotrophin, followed by Human Chorionic Gonadotrophin 48 hours later to induce super ovulation. They were then immediately placed with fertile stud males and fertilised zygotes were harvested 12 hours post-conception. CRISPR guide mRNA or plasmid DNA, and 50-100 ng/μL CRISPR-Cas9 protein were co-injected into the pronucleus by Nikki Ross and Jenna Lowe at the JCSMR Transgenesis Facility. Following injection, zygotes were incubated overnight (37 °C, 5% CO_2) until they reached the 2 cell stage. Unviable embryos were discarded, whilst viable embryos were implanted into the right uterine infundibulum of anesthetised pseudopregnant Swiss mice and left to develop until the desired dissection time point or pups were born.

2.13 Microscopy

2.13.1 Differential Interference Contrast (DIC) Microscopy

WMISH embryos were post-fixed in 4% PFA in PBS for one hour at room temperature and transferred via a glycerol series (50%, 80%) to 100% glycerol. For photography, whole embryos were flat-mounted on a glass slide under a glass coverslip and photographed in a Nikon SMZ 21500 Stereomicroscope and DS-Ri1 camera (Nikon) with DIC optics.

For node measurements, fixed embryos were mounted under a glass coverslip in 100% glycerol and their node visualized with a 40X objective (Leica) in a compound microscope (Leica DM5500 FL DIC) using DIC optics. Images were captured using a Leica DFC365 FX camera and LAS V4.3 software. X-Gal stained cells expressing *lacZ* were imaged at 40X with the same microscope.

2.13.2 Fluorescence Microscopy

Cell immunofluorescence was photographed with a 100X oil immersion objective (Zeiss) using a Zeiss Axio Observer with a Apotome Nikon TE3000 inverted fluorescence microscope. Images

were captured using a Zeiss AxioCamMRm CCD camera using DIC optics. Fluorescence was imaged at 495 nm and 590 nm (Table 2.7).

2.13.3 Scanning Electron Microscopy (SEM)

Mouse embryos at 8.0 dpc were dissected in 10% (v/v) FBS in PBS and fixed overnight in fresh 2% PFA/2.5% gluteraldehyde/0.1 M cacodylate buffer (pH 7.4) at 4 °C. After rinsing with 0.1 M cacodylate buffer, embryos were postfixed in 1% osmium tetroxide/0.1 M cacodylate for 20 minutes at room temperature. They were dehydrated through a graded EtOH series and dried at a critical point with a CPD010 (Balzers Union). Embryos were coated in platinum by an EMTECH K550X sputter coater. All imaging was performed on a Hitachi 4300SE/N FESEM at 3 kv.

2.14 Analysis

2.14.1 Node and Cilia Measurements

Kumba node morphology and size was examined using DIC optics on three embryos of each genotype (*Zic2*$^{+/+}$, *Zic2*$^{Ku/+}$ and *Zic2*$^{Ku/Ku}$). Node circumference and the length of the anterior-posterior axis were measured using LAS V4.3 software. Cilia frequency and length were counted in three embryos from each genotype (*Zic2*$^{+/+}$, *Zic2*$^{Ku/+}$ and *Zic2*$^{Ku/Ku}$) using SEM. For cilia frequency and length analysis, SEM images were recorded at 15000X magnification. The file name of each image was altered to a number and the file order randomized by an independent worker so that the genotype of the embryo was unknown when calculating cilia frequency and length, and node circumference. Cilia length was determined by measuring pixel length in Adobe Photoshop CS5 and converting it to μm (using a factor determined by the number of pixels per μm). In total, ten node cilia were measured in each of three embryos per genotype.

2.14.2 Assessing Transcript Decay Rate

To measure the rate of *ZIC2* mRNA decay, cells were either transfected with constructs containing V5-*ZIC2*-3'UTR or V5-*ZIC2*. 24 hours post transfection, the media was aspirated from cells and replaced with either supplemented DMEM (1 mL) containing 200 μM 5,6-dichloro-beta D-ribofuranosyl-benzimidazole (DRB) suspended in dimethyl sulfoxide (DMSO), or supplemented DMEM containing an equivalent volume of 0.2% DMSO as a control. Immediately, the media from a dish containing cells that had not been treated with DRB was removed and cell lysates were stored at -80 ^{o}C until samples from all time points were collected. This sample was designated as the 0 hour time point. Cells treated with DRB were subsequently lysed at 1 hour, 2 hours and 4 hours post DRB treatment. The RNA was then extracted and purified as described in Section 2.2.18.

The RNA samples were run through RT-qPCR analysis (Section 2.4) and the fold change in *ZIC2* expression in each cell sample quantified. The amount of *ZIC2* mRNA remaining after DRB treatment was calculated as a percentage of the *ZIC2* levels in untreated cells from the same time point. The proportion of *ZIC2* mRNA remaining in inhibited cells was plotted against time used to calculate the half-life ($t_{1/2}$) of *ZIC2* in Microsoft Excel using the formula:

$$t_{1/2} = \frac{\log(2)}{\log\{\mathrm{LOGEST}[(T_i, Z_i) : (T_f, Z_f)]\}}$$

The LOGEST function estimates the slope of all points over the time course using a nonlinear least squares regression (T_i is the initial time point, T_f is the final time point, Z_i is the initial proportion of *ZIC2* mRNA, Z_f is the final proportion of *ZIC2* mRNA) (Geisberg *et al.*, 2014).

2.14.3 Rapid Amplification of cDNA Ends (3' RACE)

Wildtype embryos at 7.5 dpc, 8.5 dpc and 9.5 dpc were collected and pooled based on their age and the total RNA extracted as per Section 2.2.18. A dT-adapter oligonucleotide (RA1703; Integrated DNA Technologies) was used with the SuperScipt III One-Step RT-PCR system with Platinum Taq kit (Invitrogen, Cat. No. 12574018) to convert the embryonic mRNA to cDNA, according to the manufacturer's instructions. The presence of the 'TTT' string ensures only RNA with a poly(A) tail is converted into cDNA. Newly created cDNA was purified via ammonium acetate clean up (Section 2.2.10).

A TD60 PCR was performed using an oligonucleotide 5' to the *Zic2* 3'UTR (RA98; Appendix Table A1.2) and an oligonucleotide that recognises the dT-adapter sequence (RA1704), in conjunction with 2X ImmoMix and 1 M Betaine. Following gel electrophoresis to resolve the PCR amplicon fragments, clear bands were gel extracted. If the PCR product was smeared, the entire PCR reaction was purified via ammonium acetate. The amplicons were re-PCRed with RA1704 and an oligonucleotide (RA99) nested slightly 3' to RA98. Gel electrophoresis was performed and clear amplicons gel extracted. Following purification, amplicons were TA cloned into pGEM-T-Easy and the ligation reaction electrotransformed into DH5α *E. coli*. Following bacterial colony PCR to identify transcripts predicted to be the *Zic2* 3'UTR, plasmid DNA was purified and Sanger sequenced.

2.14.4 Bioinformatics Tools

Many bioinformatics tools, outlined in Table 2.15 and Table 2.16, were employed for the analysis of the *ZIC2* NCE and related sequences. The default settings were used for each tool, unless otherwise indicated. Sequence alignment figures and sequence annotations were generated in

Geneious Pro v5.5.9. Methylation and acetylation values were analysed in Microsoft Excel to correlate enrichment with the *ZIC2* 3'UTR sequence.

2.14.5 Statistics

Luciferase assays: The standard deviation (S. D.) of the internal repeats of the representative experiment was calculated using Microsoft Excel. Pooled data was analysed using Genstat software to perform either an unpaired Student's T-Test or two-way Analysis of Variance (ANOVA) with Fischer's unprotected post hoc test on a minimum of three external repeats, and to calculate the standard error of the mean (S. E. M.).

RT-qPCR analysis: Genstat software was used perform a one-way ANOVA with Fischer's unprotected post hoc test on a minimum of two external repeats, and to calculate the S.E.M. Statistical analysis was performed on normalized C_T means.

Cilia analysis: Genstat software was used to perform a two-way ANOVA with Fischer's unprotected post hoc test on ten cilia each for three embryos per genotype, and to calculate the S.E.M.

Bioinformatics: For histone modification enrichment, values taken from Layered Tracks (ENCODE) were assigned to a nucleotide location within the *ZIC2* 3'UTR in Excel. Enrichment within the 3'UTR was compared via a one-way ANOVA with Fischer's unprotected post hoc test in Genstat software, and the S.E.M. calculated. To determine if predicted binding sites were enriched within the *ZIC2* 3'UTR, a two-tailed Chi-squared goodness of fit test was performed using the GraphPad online tool (GraphPad Software, 2017). To determine if A and U nucleotides were enriched within the *ZIC2* 3'UTR mRNA, a two-tailed Chi-squared was performed using the GraphPad online tool (GraphPad Software, 2017).

Table 2.15: Databases and software used for bioinformatics analyses.

Database/Software	Use	Reference
AREsite	Identification of AU-rich elements Poly(A) site prediction	(Gruber *et al.*, 2011)
BLASTn (NCBI)	Alignment of DNA sequences	(Altschul *et al.*, 1997)
CLUSTAL Omega	Pairwise identity percentage calculations	(Goujon *et al.*, 2010; Sievers *et al.*, 2011)
DIANA-microT-CDS	miRNA binding site prediction	(Paraskevopoulou *et al.*, 2013; Reczko *et al.*, 2012)
EMAGE	Gene expression analysis in embryos	(Richardson *et al.*, 2014)
ENCODE Genome Editor (via the UCSC Genome and Table Browser)	Identification of chromatin DNase I hypersensitivity sites Identification of chromatin methylation and acetylation enrichment Identification of p300 binding enrichment Transcription factor binding site prediction	(Kent *et al.*, 2002)
Geneious Pro 5.5.9	Alignment of DNA sequences Generating alignment figures Annotation of sequences	(Kearse *et al.*, 2012)
JASPAR	Transcription factor binding site prediction	(Mathelier *et al.*, 2014)
MGI GXD	Gene expression analysis	(Finger *et al.*, 2011)
miRanda	miRNA binding site prediction	(Betel *et al.*, 2008; John *et al.*, 2004)
miRNASNP	miRNA binding site prediction	(Gong *et al.*, 2015)

	miRNA thermodynamics binding prediction	
mrSNP	miRNA thermodynamics binding prediction	(Deveci *et al.*, 2014)
PolyA Signal Miner	Poly(A) site prediction	(Liu *et al.*, 2003)
Phylogeny.fr	Phylogenetic tree generation	(Dereeper *et al.*, 2008)
RBPDB	RNA binding protein binding site prediction	(Berglund *et al.*, 2008)
RBPmap	RNA binding protein binding site prediction	(Paz *et al.*, 2014)
TRANSFAC	Transcription factor binding site prediction	(Matys *et al.*, 2006)
UCSC Genome and Table Browser	Identification of histone modifications	(Rosenbloom *et al.*, 2015)

Table 2.16: Tracks used with the UCSC Genome Browser and ENCODE Genome Editor.

Track	**Use**	**Cell type**
CSHL Long RNA-seq	Identification and alignment of long (>200 nt) RNA transcripts	GM12878, HeLa, Hi-hESC and more
CSHL Short RNA-seq	Identification and alignment of short (<200 nt) RNA transcripts	GM12878, HeLa, Hi-hESC and more
Layered Tracks (ENCODE)	Chromatin monomethylation (H3K4Me1) enrichment Chromatin trimethylation (H3K4Me3) enrichment Chromatin acetylation (H3K27Ac) enrichment	H1-hESC
Open Chromatin DNase I HS (Duke)	Chromatin DNase I hypersensitivity sites	hESC
SC584 Standard ChIP-seq ENCODE/SYDH	p300 binding enrichment	GM12878

Chapter 3: Cardiac defects in *Zic2* mutant embryos are caused by aberrant node function

This chapter contains text from the papers: Barratt, K.S., Glanville-Jones, H.C. & Arkell, R.M., 2014. The *Zic2* gene directs the formation and function of node cilia to control cardiac situs. *Genesis*, 52(6), pp.626–635 and Diamand, K.E.M., Barratt, K.S. & Arkell, R.M (2017) Chapter 10: Overview of rodent Zic genes, "Zic family - Evolution, Development and Disease", Springer (in publication).

3.1 Introduction

Mouse embryos homozygous for the *Ku* allele of *Zic2* (*Zic2*$^{Ku/Ku}$) have apparently normal anterior primitive streak function with molecular abnormalities first detected at the stage of overt node formation (7.0 dpc). The expression of any gene so far examined that would normally be present in the 7.0 dpc node or its derivatives is greatly diminished in *Zic2*$^{Ku/Ku}$ embryos (Warr *et al.*, 2008). Tissues derived from the later stage node (such as the trunk notochord) are more mildly affected in 9.5 dpc embryos (see Fig.3 O-R of Warr et al. 2008) suggesting that the defect in node gene expression and function is transient. Given that ZIC2 is required to execute the A-P and D-V components of node function, it is hypothesised that the node and midline formation defects in *Zic2*$^{Ku/Ku}$ embryos would also interfere with L-R axis formation. This is supported by the observations that some homozygous embryos have incorrect heart morphology at 9.5 dpc (see Fig.3 O,P of Warr et al. 2008) and that approximately 5% of human *ZIC2*-associated HPE probands exhibit co-morbidities such as cardiac abnormalities (Solomon *et al.*, 2010a).

The establishment of the L-R embryonic axis occurs at ~8.0 dpc in the mouse, when a leftward flow of fluid can be detected in the mouse embryonic node (Field *et al.*, 2011; Norris, 2012). To create this axis, four main steps must occur to achieve the correct orientation and asymmetry of organs: the breaking of embryonic symmetry, the establishment of leftward nodal flow, asymmetric gene expression and the nodal cascade, and the placement of developing organs. It is the failure of these steps to correctly pattern the embryo that results in congenital heart defects and L-R axis defects such as Heterotaxy.

3.1.1 Morphogenesis of the murine node

The murine node forms as a pit shaped structure at the anterior of the primitive streak at mid-gastrulation, flanked by endoderm and the paraxial mesoderm, and flanked again by the lateral plate mesoderm (LPM) which contributes to asymmetric organ structure after axis

establishment (Norris, 2012) (Figure 3.1). At the mid-streak stage, the outer surface of the embryo is covered by squamous endoderm cells, with APS cells arranged as a sheet beneath the endoderm layer (Yamanaka *et al.*, 2007). Lineage tracing found that these APS cells begin to express *Noto*, a homeobox transcription factor, and organise into columnar epithelium with ventral facing apical surfaces, distinguishing them from other APS cell populations (Yamanaka *et al.*, 2007). These putative node cells then begin to delaminate from the APS. By the late-streak stage (7.0 dpc), several clusters of 200-300 columnar epithelial cells, called pit cells, with small apical surfaces and cilia have emerged between the endoderm cells and can be seen into a shallow, crescent shaped depression on the ventral side of the embryo (Sulik *et al.*, 1994; Yamanaka *et al.*, 2007). Approximately 25-30 crown cells surround the node, extending higher into the extracellular space than the pit cells to create a barrier around the node (Figure 3.1) (Bellomo *et al.*, 1996; Nonaka *et al.*, 1998; Norris, 2012; Sulik *et al.*, 1994; Yamanaka *et al.*, 2007). By the time the node is completely revealed, the pit and crown cells are contiguous with the surrounding DE, where they transit the node and migrate to the anterior midline, as described in Chapter 1 (Kinder *et al.*, 2001; Robb and Tam, 2004). As cells move out from the node to form the notochord, cells from the APS are continuously recruited to repopulate the node (Kinder *et al.*, 2001).

Each pit and crown cell in the ventral layer of the node carries a monocilium on its apical surface that extends into the extracellular space (Lee and Anderson, 2008; Sulik *et al.*, 1994). Cilia on the pit cells rotate in a clockwise direction and support a flow of extracellular fluid towards the left side of the node (called nodal flow) that is posited to establish differential signal(s) on the left and right of the node (Nonaka *et al.*, 2002, 1998). The cilia on the crown cells are generally immotile, and these sensory cilia are thought to interpret the signal(s) established via nodal flow (McGrath *et al.*, 2003). Whilst the location and distribution of cilia in the node pit has been shown to have no effect on the ability to generate a leftward nodal flow, the angle at which the cilia project from the pit cells is vital. Wildtype cilia tilt posteriorly and rotate clockwise at ~600 rpm (Nonaka *et al.*, 2005; Yoshiba and Hamada, 2014). Studies of defective cilia in the mouse, ESCs, and zebrafish have shown that when the cilia do not tilt at the correct angle, or a subset tilt anteriorly instead of posteriorly, a vortex is created above the cilia that results in randomisation of nodal flow (Maisonneuve *et al.*, 2009; Okada *et al.*, 2005, 1999). It is disruptions such as these in the formation or function of node cilia that lead to the loss of unidirectional flow of fluid across the node and result in L-R patterning defects (Yoshiba and Hamada, 2014).

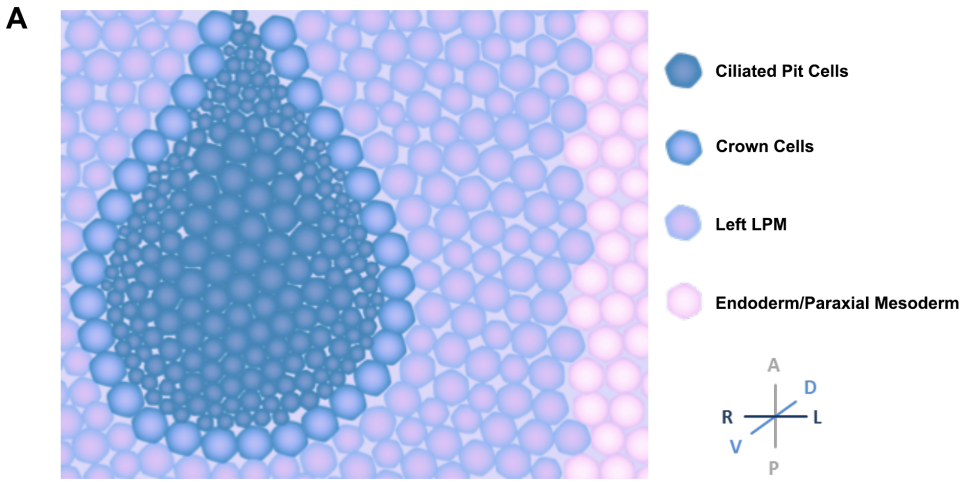

A
Ciliated Pit Cells
Crown Cells
Left LPM
Endoderm/Paraxial Mesoderm
A
D
R
L
V
P
B
1
2
3
4
5
Notochord
Lateral plate mesoderm
Pit cells
Node
Crown cells
Emerging posterior mesoderm
Ciliogenesis
Symmetric expression
Midline barrier formation
Nodal flow
Signal transfer
Asymmetric expression
Nodal cascade
7.0 dpc
7.5 dpc
7.75 dpc
8.0 dpc
8.25 dpc
Left-right axis formation

Figure 3.1: Schematic of L-R axis formation in the murine embryo. (a) The embryonic node contains ciliated pit cells surrounded by ciliated crown cells. Endoderm and paraxial mesoderm cells flank the node, with mesoderm forming the right (not shown) and left lateral plate mesoderm (LPM) down the sides of the embryo. **(b)** Steps involved in L-R axis formation. 1: The node is induced to form at mid-gastrulation (7.0 dpc) and in the next 24 hours develops into a shallow, crescent shaped depression on the ventral side of the embryo. 2: A monocilium, extending into the extracellular space, forms on the apical surface of each pit and crown cell; these become posteriorly polarised over time. Signalling molecules (such as NODAL) are expressed in the crown cells. 3: The cilia of the pit cells rotate in a clockwise direction directing first a disorganised, then laminar, leftward flow of extracellular fluid within the node. It is posited that this leftward nodal-flow is sensed by crown cell cilia, prompting a Ca^{2+} flux in the left crown cells which modifies gene expression. By the end of this 24 hour period (at 8.0 dpc) the first known asymmetries in gene expression are detected within node crown cells (*Dand5*, and soon thereafter *Nodal*, become asymmetrically expressed in the node crown) (reviewed in Norris 2012). WNT ligand is also asymmetrically expressed and canonical signalling then amplifies the initial *Dand5* asymmetry (Nakamura *et al.*, 2012). 4: The asymmetric signal(s) are propagated to the LPM and prevented from spreading to the right LPM by the recently formed midline barrier. 5: NODAL signalling in the left LPM controls its own expression, and that of other molecules, that ultimately direct organ position and other asymmetries (reviewed in Norris 2012). A: anterior, P: posterior, D: dorsal, V: ventral, R: right, L: left. Figure B was obtained from R. Arkell.

3.1.2 Requirements for cilia formation and function

A combination of three homeobox transcription factors, expressed prior to the generation of nodal flow, act upstream of intraflagellar transport (IFT) motor proteins and are required for the formation of immotile and motile cilia in the node. The forkhead homeobox, FOXJ1, in combination with members of the Not homeobox (NOTO) and the regulatory factor X (RFX) families (Beckers *et al.*, 2007; Didon *et al.*, 2013; Thomas *et al.*, 2010), control ciliogenesis in the embryonic node. NOTO, acting upstream of FOXJ1 and RFX3, regulates expression of components involved in cilia axonemal assembly and function (Beckers *et al.*, 2007). In contrast, RFX3 regulates growth of cilia and FOXJ1 is responsible for the differentiation of motile ciliated cells (Bonnafe *et al.*, 2004). Failure of any of the three TFs results in abnormal L-R axis development in mice. A threshold amount of FOXJ1 is required to induce expression of cilia component genes in ciliated human airway epithelium cells such as DNAL1 and SPAG6. Whilst FOXJ1 can induce cilia differentiation on its own, the addition of RFX3 as a co-factor improves this induction 2.9 fold (Didon *et al.*, 2013). FOXJ1 therefore acts both upstream of RFX3, and as a co-factor of RFX3, to co-regulate genes involved in cilia development and function (Didon *et al.*, 2013). Together FOXJ1 and RFX3 induce basal progenitor cells to differentiate into a multiciliated cell lineage.

The detection of the leftward flow across the node requires the concerted efforts of polycystins PKD2 and PKD1L1. PKD2 is ubiquitously localised to immotile cilia in the crown cells, not just restricted to the cilia on the left of the node, and is thought to be responsible for detecting the Ca^{2+} spike. This ubiquitous localisation around the entire node is perhaps due to the asymmetrical expression of *Dand5* (formally *Cerl2*) on the right of the node, which is a downstream target of flow-induced PKD2 signals (Yoshiba *et al.*, 2012). PKD2, acting as an effector, forms a complex with PKD1L1, acting as a sensor, to create a flow-sensing complex not just in the embryonic node but also in the kidney (Norris, 2012; Pennekamp *et al.*, 2002). The actual mechanisms of how this PKD2/PKD1L1 complex senses the fluid flow is yet to be identified, with suggestions of chemosensation and mechanosensation currently being investigated (Norris, 2012). Once the PKD2/PKD1L1 complex detects leftward flow, *Dand5* expression is repressed in the left, but not the right of the embryonic node. This in turn results in increase expression of *Nodal* in the left of the node (Yoshiba *et al.*, 2012) and the onset of the L-R axis formation.

3.1.3 Nodal flow

Sufficient nodal flow can be created with as little as two rotating cilia in the node of mutant embryos (Shinohara *et al.*, 2012), however the exact mechanisms behind how the flow is detected in the embryo is still a contentious issue (Norris, 2012). Two distinct populations of cilia

in the node are responsible for both the generation and the sensing of nodal flow: monocilia in the pit cells create the leftward flow and immotile sensory cilia in the crown cells detect the leftward flow (Field *et al.*, 2011). It has been hypothesised that, as a result of nodal flow, certain molecules become enriched on the left side of the node. Detection of this higher concentration leads to an asymmetric Nodal cascade. The exact nature of the morphogen(s) required for this process is still unknown, but this theory has been shown to be plausible via computational models providing the molecules in question are between 15 and 50 kDa. In response to nodal flow, a Ca^{2+} concentration spike can be detected at the left of the node but not the right (Norris, 2012). When Ca^{2+} blockers such as $GdCl_3$ and 2-ABP are introduced, asymmetric gene expression in crown cells is disrupted (Yoshiba *et al.*, 2012), suggesting Ca^{2+} is one of the main morphogens involved in establishing the L-R axis. This elevation of Ca^{2+} is thought to come about from signals packaged in nodal vesicular particles, which in turn are regulated by SHH, FGF and retinoic acid morphogens (Hirokawa *et al.*, 2006).

3.1.4 The Nodal cascade

Dand5 is the first known asymmetric gene expressed in the node, being detected in crown cells at 8.0 dpc. *Nodal* expression follows shortly after. Whilst *Nodal* exhibits a left sided bias in the crown cells, its antagonist *Dand5* is expressed with a right sided bias (Collignon *et al.*, 1996; Lowe *et al.*, 1996; Marques *et al.*, 2004; Norris and Grimes, 2012; Pearce *et al.*, 1999). It is during this time that Ca^{2+} becomes elevated at the left boundary of the node (McGrath *et al.*, 2003). Subsequently, asymmetric signals are propagated from the node to the left LPM via an unresolved mechanism. By the 3-somite stage of development (8.25 dpc), *Nodal* signalling in the left LPM has induced its own expression, as well as that of the *Nodal* antagonist *Lefty2,* along with the TF *Pitx2* in an event known as the Nodal cascade (Shiratori and Hamada, 2006). Activation of the Nodal cascade coincides with cilia establishing a faster nodal flow than before (Norris, 2012; Yoshiba *et al.*, 2012). The expression of another NODAL antagonist, *Lefty1*, at the embryonic midline forms a barrier that prevents the spread of the Nodal cascade to the right LPM (Meno *et al.*, 1998), fortifying the newly established L-R axis. When *Lefty1* expression is disrupted, this barrier breaks down and asymmetry occurs (Meno *et al.*, 1998).

Nodal and *Lefty2* are only transiently expressed in the left LPM for ~8 hours , in contrast to *Pitx2* which, once activated, remains asymmetrically expressed in the LPM for two days during organogenesis (Ryan *et al.*, 1998; Yoshioka *et al.*, 1998). For this left-sided expression to occur, the asymmetrical signals generated in the node must be transferred to the left. Whilst the best candidate for transfer is currently a NODAL protein, studies have yet to confirm if this is indeed the molecule that initiates left LPM asymmetric expression (Yoshiba *et al.*, 2012). Research has instead focused on the route by which the signal is transferred from the node to the LPM. Recent

studies have suggested that Ca^{2+} signals are transferred from the node to the LPM via gap junctions in the definitive endoderm, and it is these signals that contribute to the activation of the Nodal cascade (Viotti *et al.*, 2012). When SOX17, a homeobox TF required for definitive endoderm development, is absent, embryos exhibit L-R patterning defects, a defective Nodal cascade, and are unable to migrate a dye between endodermal cells, signalling the loss of gap junction activity (Viotti *et al.*, 2012). In comparison to the LPM and definitive endoderm, the midline barrier between the left and the right of wildtype embryos contains no gap junctions, thus preventing signals from crossing to the right of the embryo (Viotti *et al.*, 2012).

3.1.5 Heterotaxy and congenital heart malformations

Vertebrate organ asymmetry is highly conserved (Norris, 2012), indicating a fundamental function for the placement and orientation of each organ. Though defects in organ *situs* are relatively rare in humans, they are most commonly associated with congenital heart disease and ciliopathies (Norris, 2012). Alterations in the L-R axis are defined as *situs inversus*, whereby the organs are arranged as the complete mirror image compared to an unaffected person, or heterotaxy (*situs ambiguous* and isomerisms), whereby there is a lack of concordance in the placement of organs, or mirror image duplication of paired organs (such as two lungs that both contain three lobes). Though pure *situs inversus* is not commonly associated with intracardiac cases, heterotaxy accounts for 3% of congenital heart defects (Brueckner, 2007; Zhu *et al.*, 2006). Additionally, defects in the L-R patterning have been linked to a range of other situs abnormalities such as biliary atresia, intestinal malrotations, stomach curvature, vertebral anomalies and respiratory distress from primary ciliary dyskinesia (Cohen, 2012; Davis *et al.*, 2017; Zhu *et al.*, 2006), as well as alteration in the directionality of embryo turning (Collignon *et al.*, 1996). All of these alterations in organ placements can be traced back to defects in the establishment of the L-R axis during embryogenesis, with the direction of heart looping and the directionality of other organs appearing to be independent processes in the same pathway (Chin *et al.*, 2000; Harvey, 2002).

3.1.6 Chapter 3 aims

I set out to quantify the extent of the heart defects seen in *Zic2*$^{Ku/Ku}$ embryos, and sought to determine if a defective node was the cause.

3.2 Results

3.2.1 Cardiac situs is randomised in *Zic2*$^{Ku/Ku}$ embryos

To determine whether the previously characterised defect in midline development in *Zic2*$^{Ku/Ku}$ embryos impacts formation of the L-R embryonic axis, the direction of heart looping was scored in 9.5 dpc embryos (N=14 *Zic2*$^{+/+}$, N=41 *Zic2*$^{Ku/+}$ and N=16 *Zic2*$^{Ku/Ku}$) WMISH to *Nppa*. As *Nppa* is expressed in the right atrium and left ventricle of the developing heart (Moorman and Christoffels, 2003), it was used to confirm the identity of heart regions. All wildtype, heterozygous and 56% of *Zic2*$^{Ku/Ku}$ embryos presented with dextral looping (Figure 3.2A,D), whereas 31% of *Zic2*$^{Ku/Ku}$ embryos exhibited a leftward curve of the heart tube (sinistral looping, Figure 3.2B,E) and the remaining 13% of *Zic2*$^{Ku/Ku}$ embryos had a heart tube that looped forwards (ventral looping, Figure 3.2C,F).

3.2.2 *Zic2*$^{Ku/Ku}$ embryos exhibit additional asymmetries

The mid-gestation demise of *Zic2*$^{Ku/Ku}$ embryos (Elms *et al.*, 2003) prevented the direct scoring of later asymmetric organs. Stereotypic cardinal vein morphology, however, occurs in response to L-R axis establishment and can be independent of heart looping. For example, defects in cardinal vein morphology occur in conjunction with other organ asymmetries such as lung isomerisms and mis-location of the pancreas in embryos that lack *Pitx2* (Shiratori *et al.*, 2006). *Pitx2* is expressed in the left cardinal vein (Meno *et al.*, 1998) and was used as a surrogate marker for the correct establishment of asymmetries other than heart looping. At 9 – 9.5 dpc, *Pitx2* is expressed in the cardinal vein on the left, but not right, of wildtype embryos (Figure 3.3A,C). Left cardinal vein expression of *Pitx2* was detected in the majority of *Zic2*$^{Ku/Ku}$ embryos, but some embryos lacked this expression domain, independent of the direction of heart looping (Figure 3.3B). Other embryos exhibited looping defects but maintained normal cardinal vein *Pitx2* expression (Figure 3.3D). This suggests that aspects of L-R organ development other than heart looping may be affected in *Zic2*$^{Ku/Ku}$ embryos. Interestingly, ectopic expression of *Pitx2* could be seen in the midbrain of 9.5 dpc *Zic2*$^{Ku/Ku}$ embryos (Figure 3.3D). *Pitx2* is expressed in the ventral diencephalon (Martin *et al.*, 2002), but not until 10.5 dpc, and wildtype stage-matched embryos did not have corresponding *Pitx2* expression.

3.2.3 The Nodal cascade is absent in *Zic2*$^{Ku/Ku}$ embryos

Overt L-R axis formation is preceded by the asymmetric expression of certain genes in the LPM and by the establishment of a midline barrier. At the 3-6 somite stage of development the secreted molecule *Nodal*, its secreted antagonist *Lefty2* and the downstream transcription factor *Pitx2* are all expressed in the left LPM and the Nodal antagonist *Lefty1* is expressed at the midline. Examination of the expression of these genes in early somite stage embryos via

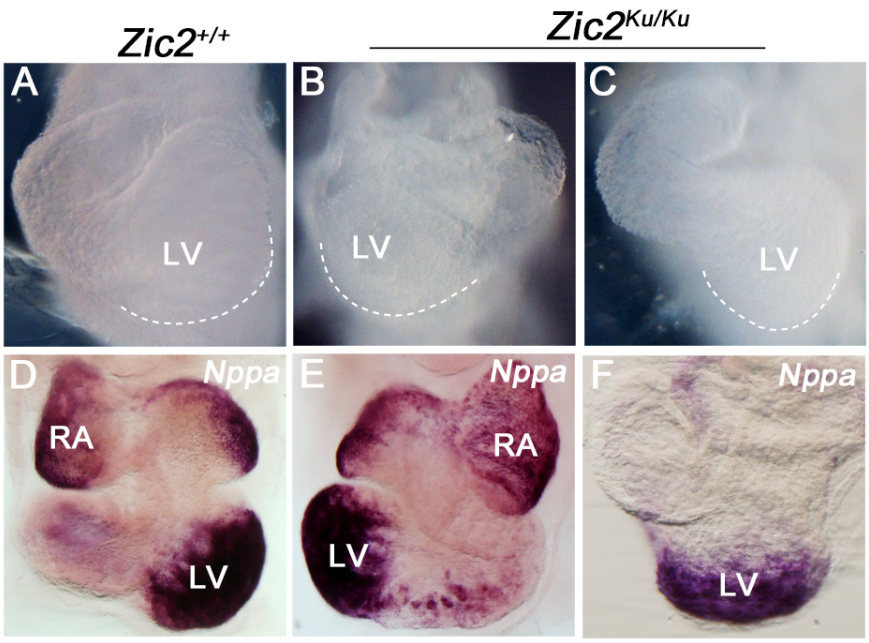


Figure 3.2: *Zic2*$^{Ku/Ku}$ embryos exhibit cardiac situs defects. (a-c) Dark-field images of hearts from embryos with 21 somites in ventral view; the dotted line shows the extent of the heart tube. **(a)** Wildtype heart formation (dextral looping). **(b)** A *Zic2*$^{Ku/Ku}$ embryo with reverse heart formation (sinistral looping). **(c)** A *Zic2*$^{Ku/Ku}$ embryo with abnormal forward looping (ventral looping). **(d-f)** Images of hearts following WMISH to *Nppa* to confirm chamber identity. **(d)** A 25-somite *Zic2*$^{+/+}$ embryo with dextral looping. **(e)** A 25 somite *Zic2*$^{Ku/Ku}$ embryo with sinistral looping. **(f)** A 16 somite *Zic2*$^{Ku/Ku}$ embryo with ventral looping. LV: left ventricle, RA: right atrium. Dotted line indicates heart looping.

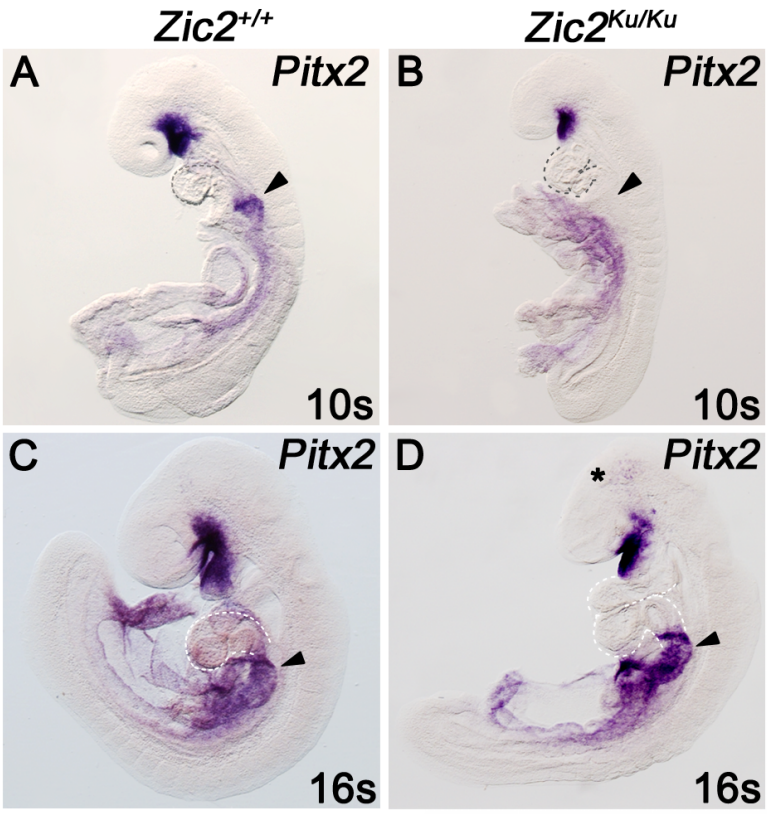


Figure 3.3: Cardinal vein situs is randomized in *Zic2*$^{Ku/Ku}$ embryos. Lateral view of embryos following WMISH to *Pitx2*. **(a)** A 10 somite *Zic2*$^{+/+}$ embryo with *Pitx2* transcription in the left cardinal vein (arrowhead). **(b)** A 10 somite *Zic2*$^{Ku/Ku}$ embryo lacking *Pitx2* cardinal vein expression (arrowhead). **(c)** A 16 somite *Zic2*$^{+/+}$ embryo with *Pitx2* transcription in the left cardinal vein (arrowhead). **(d)** A 16 somite *Zic2*$^{Ku/Ku}$ embryo with sinistral heart looping, *Pitx2* transcription in the left cardinal vein (arrowhead) and ectopic forebrain expression (asterisk). Dotted line indicates heart looping.

WMISH demonstrated that each consequence of Nodal signalling is severely compromised in *Zic2*$^{Ku/Ku}$ embryos. The expression of *Nodal* was detected at the node but not in the LPM of 3-somite mutant embryos (Figure 3.4A, B). Likewise, *Lefty2* expression was not detected in the LPM (Figure 3.4C, D). The expression of *Lefty1* at the embryonic midline was severely depleted in early somite-stage *Zic2*$^{Ku/Ku}$ embryos (Figure 3.4C, D). The absence of the midline barrier in the presence of the asymmetric signal is typically associated with the bi-lateral establishment of the Nodal cascade (i.e. bi-lateral expression of *Nodal*, *Lefty2* and *Pitx2*) (Meno *et al.*, 1998; Yamamoto *et al.*, 2003).

The expression of *Pitx2* was detected throughout the entire A-P extent of the LPM in wildtype embryos with 3-5 somites, whereas it was only weakly detected in the anterior LPM of equivalent stage *Zic2*$^{Ku/Ku}$ embryos (Figure 3.4E, F). The apparent ability to initiate anterior LPM *Pitx2* expression in the absence of *Nodal* expression suggests loss of *Zic2* function reveals a Nodal independent mechanism of *Pitx2* expression initiation. The near absence of both the midline barrier and LPM expression of *Nodal*, *Lefty2* and *Pitx2* suggests that the asymmetric signal is generated and perceived but not transferred to the LPM, or that it is not correctly established.

3.2.4 Definitive endoderm gap-junction function is normal in *Zic2*$^{Ku/Ku}$ embryos

The correct formation of the definitive endoderm is required to transfer the asymmetric signal(s) to the LPM via gap-junctions, with embryos lacking *Sox17* expression unable to complete this process (Viotti *et al.*, 2012). The definitive endoderm of *Zic2*$^{Ku/Ku}$ embryos is aberrant at mid-late gastrulation (Warr *et al.*, 2008) and may be incompetent for signal transfer. WMISH to *Sox17* revealed no difference in expression between wildtype and mutant embryos (Figure 3.5), suggesting the definitive endoderm is capable of signal transmission between the node and left LPM.

3.2.5 Node function and cilia development is compromised in *Zic2* mutant embryos

To determine whether the asymmetric midline signal is effectively established, gene expression at the node of early somite embryos was examined in stage-matched wildtype and mutant embryos. The expression of *Nodal* itself becomes asymmetric at the node of wildtype embryos slightly before the initiation of *Nodal* LPM expression (i.e. at the 0-2 somite stage). The perinodal expression domain of *Nodal* varied between homozygous *Ku* embryos, but was always different to that of stage-matched wildtype littermates (Figure 3.6A-D). In all cases *Nodal* expression was diminished and in some cases expression failed to become asymmetric. Expression of another secreted molecule, the NODAL antagonist *Dand5*, also becomes asymmetric at the node of wildtype embryos prior to the initiation of *Nodal* LPM expression.

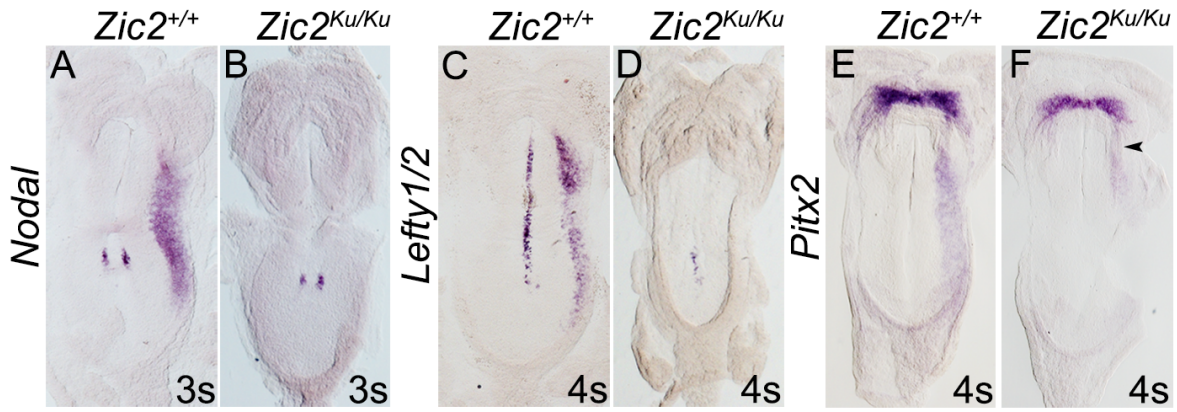


Figure 3.4: The Nodal cascade and midline barrier are compromised in *Zic2*$^{Ku/Ku}$ embryos. Ventral views of embryos following WMISH to the genes shown on embryos of the genotypes and stage shown. Anterior is to the top in all images. **(a, b)** *Nodal* is normally expressed in the crown cells of the node and in the left LPM; the LPM expression is absent in *Zic2*$^{Ku/Ku}$ embryos. **(c, d)** *Lefty1* is normally expressed in the midline rostral of the node and *Lefty2* in the left LPM; *Lefty1* expression is depleted and *Lefty2* expression lost in *Zic2*$^{Ku/Ku}$ embryos. **(e, f)** *Pitx2* is normally expressed bilaterally in the cranial mesenchyme and in the left LPM; the cranial mesenchyme expression is retained and residual expression is seen in the anterior left LPM of *Zic2*$^{Ku/Ku}$ embryos (arrowhead in F). S: somites.

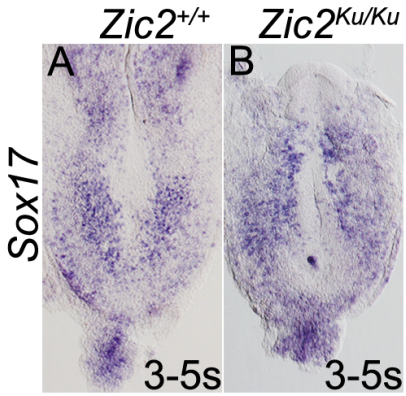


Figure 3.5: The definitive endoderm is normal in *Zic2*$^{Ku/Ku}$ embryos. Ventral views of embryos following WMISH to *Sox17* of the genotypes and stage shown. Anterior is to the top in all images, only the caudal half of the embryo is shown. **(a)** *Sox17* is normally expressed in the definitive endoderm; **(b)** this expression is retained in *Zic2*$^{Ku/Ku}$ embryos. S: somites.

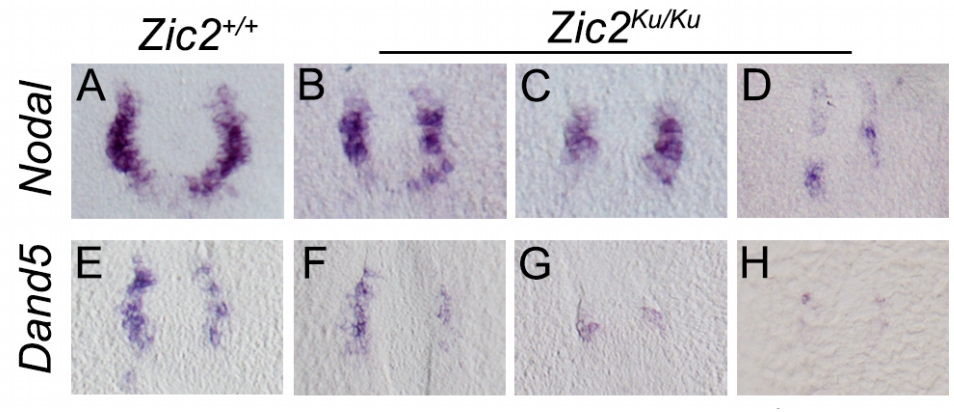


Figure 3.6: Aberrant gene expression in the early somite node of *Zic2*$^{Ku/Ku}$ embryos. Ventral view of 3-5 somite embryos of the genotypes shown following WMISH to the genes shown. Anterior is to the top. **(a-d)** Nodal expression normally becomes asymmetric in wildtype 2 somite embryos; the peri-nodal expression is diminished and/or fails to become asymmetric in *Zic2*$^{Ku/Ku}$ embryos. **(e-h)** *Dand5* is asymmetrically expressed in the crown cells of the node in wildtype 2 somite embryos; the peri-nodal expression is diminished and fails to become asymmetric in *Zic2*$^{Ku/Ku}$ embryos.

The expression of *Dand5* was greatly depleted in the node of all *Zic2*$^{Ku/Ku}$ embryos examined and no asymmetric expression was observed (Figure 3.6E-H). These data indicate that *Zic2* is genetically upstream of asymmetric gene expression at the node.

To determine whether the effects on gene expression at the node result from aberrant node development we examined the node of 3-5 somite-stage embryos by light microscopy (DIC optics) and the node of 0-3 somite-stage embryos by SEM. Visual inspection of the overall morphology of the node (shape and size) suggested that *Zic2*$^{Ku/Ku}$ nodes were smaller (Figure 3.7A-D), however measurement of the node circumference and length of the anterior-posterior axis were not found to be significantly different from wildtype nodes (p>0.05, Figure 3.8A-B). The cilia of node pit cells occurred at the same frequency (i.e. approximately 1 cilium/cell) across all genotypes. Cilia length was overtly different (Figure 3.7E-F') and when measured wildtype embryos were found to contain cilia with a mean length of 4 µm. In contrast, *Zic2*$^{Ku/+}$ embryos had cilia with a mean length of 3.1 µm whilst *Zic2*$^{Ku/Ku}$ embryos had cilia with a mean length of 2.5 µm (p<0.01) (Figure 3.8C). In addition, the cilia of *Zic2*$^{Ku/Ku}$ embryos were dysmorphic and often bulbous at the base (Figure 3.7F-F'). In combination with the gene expression data, this suggests that the node cilia do not retain sufficient function for symmetry breaking.

3.2.6 ZIC2 is required for the correct expression of genes involved in cilia formation and function

Zic2 is not expressed in the node of early somite stage embryos, whereas it is expressed in the node of mid-late gastrula embryos (Elms *et al.*, 2004). Since ZIC2 is a transcriptional regulator and expected to act cell-autonomously, we hypothesised that ZIC2 is required to (either directly or indirectly) regulate the expression of genes required for node and cilia formation and function in the gastrula. The Arkell lab has previously shown that the expression of *Foxa2* in the mid-gastrula node is severely depleted (Warr *et al.*, 2008) and it is known that embryos lacking FOXA2 are unable to form a functional node (Ang and Rossant, 1994). To determine whether this early defect in node function may also influence L-R axis formation, I examined the expression of genes required for node ciliogenesis (*Noto, Rfx3, Foxj1*) (Beckers *et al.*, 2007; Bonnafe *et al.*, 2004; M. Zhang *et al.*, 2004) and sensory function (*Pkd1l1*) (Field *et al.*, 2011). The expression of each of these genes is severely depleted in 7.0 and 7.5 dpc embryos (Figure 3.9), suggesting widespread dysgenesis of the node of mid-gastrula *Zic2*$^{Ku/Ku}$ embryos.

3.2.7 ZIC2-wt and ZIC2-*Kumba* are not localised to cilia

The ZIC proteins are members of the Gli superfamily of transcriptional regulators (Brewster *et al.*, 1998; Mizugishi *et al.*, 2001; Sakai-Kato *et al.*, 2008). As other members (GLI and GLIS) of the

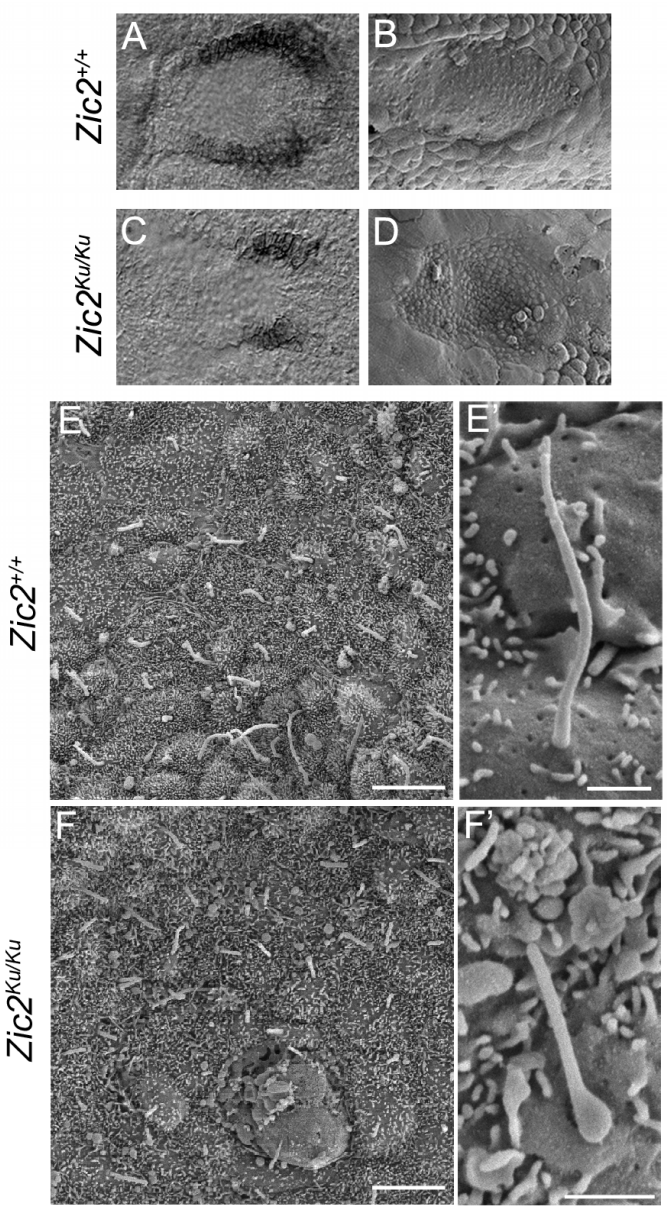


Figure 3.7: Aberrant cilia morphology in the early somite node of *Zic2*$^{Ku/Ku}$ embryos. Ventral view of 3-5 somite embryos of the genotypes shown. Anterior is to the left. **(a, c)** DIC and **(b, d)** SEM images of node morphology in embryos of the genotypes shown. The length of *Zic2*$^{Ku/Ku}$ nodes is not distinguishable from wildtype. **(e-f')** SEM images of node cilia in 3-4 somite embryos of the genotypes shown. Scale bar: 5 µm (f, g) and 1 µm (f', g').

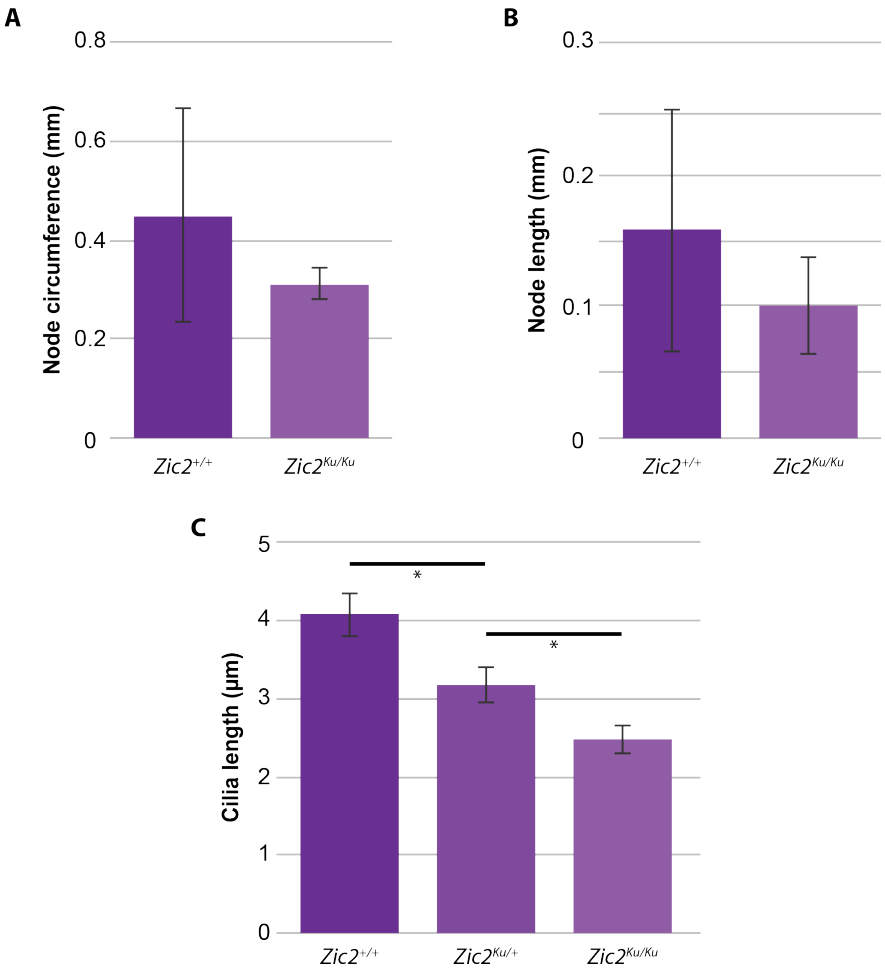


Figure 3.8: Cilia in the early somite node of *Zic2*$^{Ku/Ku}$ embryos are significantly shorter than wildtype cilia. (a-b) A column graph depicting the mean node circumference and length (mm) in 3-5 somite embryos of the genotypes shown. Error bars: SEM, *: $p<0.05$, Student's t-test. **(c)** A column graph depicting the mean cilia length (μm) in 0-3 somite embryos of the genotypes shown. Error bars: SEM, *: $p<0.01$, Fisher's LSD ANOVA.

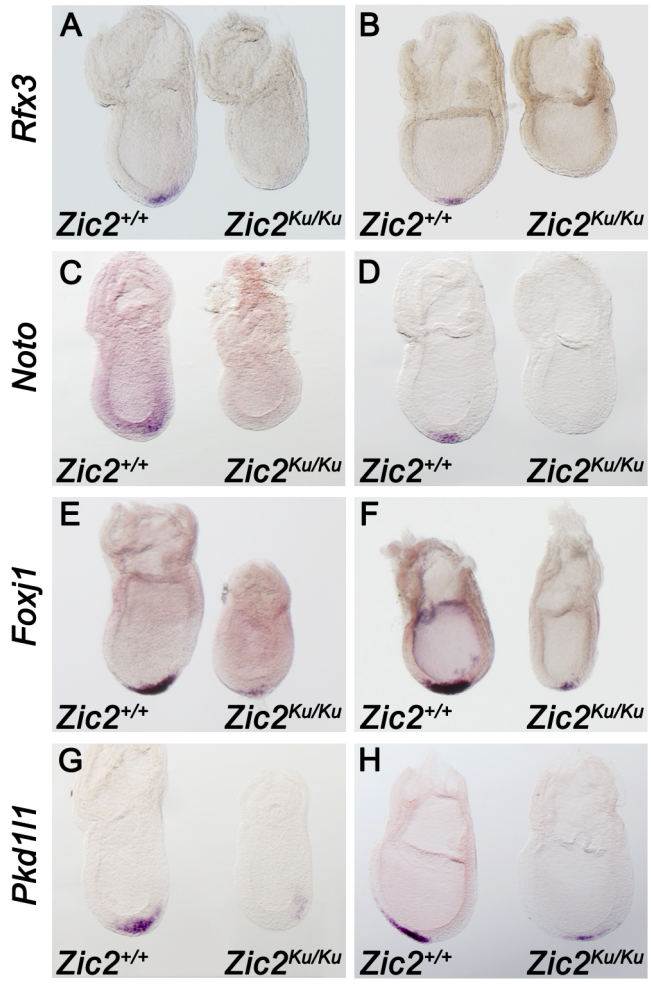


Figure 3.9: Aberrant gene expression in the mid-gastrula node of *Zic2*$^{Ku/Ku}$ embryos. Lateral views of embryos following WMISH to the genes shown on embryos of the genotypes shown. Anterior is to the left. **(a, c, e, g)** 7.0 dpc, pre-allantoic bud stage. **(b, d, f)** 7.5 dpc, early-mid bud stages, **(h)** 7.75 dpc, late bud stages. **(a-b)** *Rfx3* node expression is diminished in *Zic2*$^{Ku/Ku}$ embryos. **(c-d)** *Noto* node expression is diminished in *Zic2*$^{Ku/Ku}$ embryos. **(e-f)** *Foxj1* node expression is diminished in the *Zic2*$^{Ku/Ku}$ embryos. **(g-h)** *Pkd1l1* node expression is diminished in the *Zic2*$^{Ku/Ku}$ embryos.

superfamily are known to localise to cilium (Haycraft *et al.*, 2005; Kang *et al.*, 2010, 2009), it was hypothesised that ZIC2 would also. A ciliated cell line (LLCPK1) was used to investigate the localisation of wildtype ZIC2 and *Kumba* variant protein. Cells were transfected with V5-ZIC2-wt, V5-ZIC2-*Kumba* or PKD2-Myc as a positive control and serum starved to induce cilia formation. 24 hours post-transfection, cells were immunostained with α-V5 or α-C-Myc to identify ZIC2 and PKD2 proteins, and counterstained with α-Acetylated Tubulin to identify cilia. In comparison to PKD2-Myc, which localised in cilia, neither V5-ZIC2-wt nor V5-ZIC2-*Kumba* could be detected localised to cilia (Figure 3.10). All three proteins studied were localised in the nucleus.

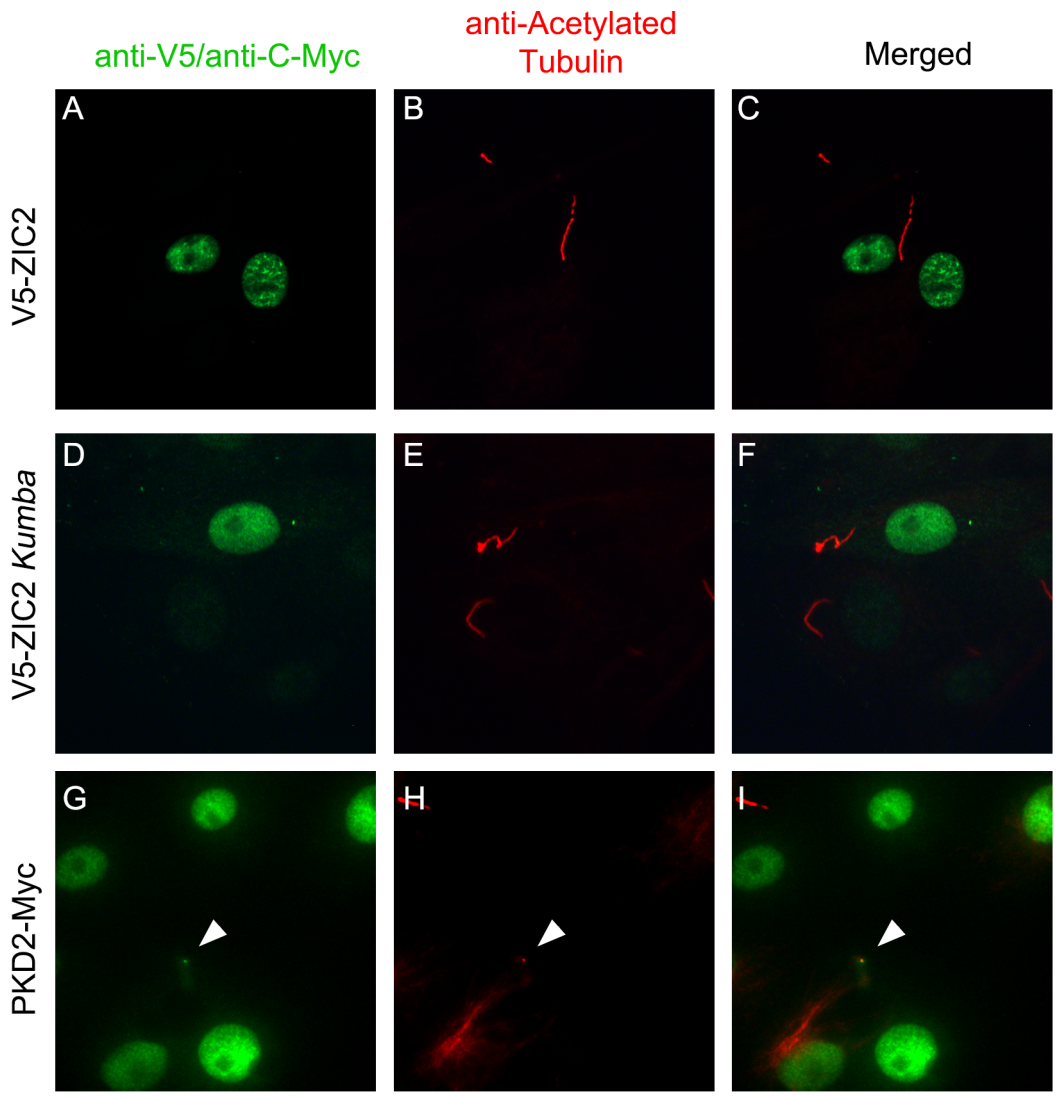


Figure 3.10: ZIC2-wt and ZIC2-*Kumba* are not localised to cilia. LLCPK1 cells were transfected with **(a-c)** V5-ZIC2-wt, **(d-f)** V5-ZIC2-*kumba* or **(g-i)** PKD2-Myc and immunostained with α-V5 (green; a, c, d and f), α-C-Myc (green; g, i) or α-Acetylated Tubulin (red; b, c, e, f, h and i). **(a-f)** V5-ZIC2-wt and V5-ZIC2-*Kumba* were not localised to the cilia. In contrast, **(g-i)** PKD2-myc is localised to the cilia (white arrows).

3.3 Discussion

It has previously been shown that ZIC2 dependent transcription is required for correct function of the 7.0 dpc node (Brown *et al.*, 2005; Warr *et al.*, 2008). In this chapter, I show that early defects in node function of *Zic2* mutants also compromise the establishment of the L-R axis. ZIC2 evidently acts upstream of the expression of genes known to regulate ciliogenesis and function, since the expression of these genes within the node is greatly depleted and ZIC2 does not appear to localise in the cilia. Pit cells within the node of embryos that lack ZIC2 function have short, dysmorphic cilia relative to their wildtype, stage-matched littermates and the molecular hallmarks of symmetry breaking at the embryonic midline are abnormal. Subsequently, the Nodal signalling cascade within the left LPM fails and cardiac situs is randomised.

3.3.1 Signalling interactions in the embryonic node and cilia defects

Zic2 loss-of-function leads to L-R defects most likely because it controls the expression of genes required for the formation and function of node cilia. Each of the other transcription factors known to be involved in this process (*Noto*, *Foxj1* and *Rfx3*) uniquely affects the L-R axis formation component of node function. In contrast ZIC2 appears to act upstream of genes required for axial mesoderm as well as cilia formation and so influences the development of all three embryonic axes. Whether ZIC2 directly regulates these ciliogenesis genes, or is upstream of genes that interact with *Noto*, *Foxj1* or *Rfx3,* is yet to be determined.

Embryos that lack *Noto*, *Foxj1* or *Rfx3* have disrupted L-R axis formation and presumably have compromised nodal flow due to cilia malformation, but they are often able to initiate the Nodal cascade within the LPM (with cases of left, right or bi-lateral expression found in combination with a proportion of embryos that fail to show LPM expression of the relevant marker genes) (Beckers *et al.*, 2007; Bonnafe *et al.*, 2004; M. Zhang *et al.*, 2004). A combination of mutagenesis and transcriptional profiling experiments in a variety of model organisms has led to the conclusion that FOXJ1 is necessary for the biogenesis of motile cilia, whereas RFX proteins are necessary for assembly of both motile and immotile cilia (Thomas *et al.*, 2010).

In contrast to $Zic2^{Ku/Ku}$ embryos, in which indicators of the Nodal cascade in both the LPM and midline were almost entirely absent and neither right-sided nor bi-lateral expression of *Nodal*, *Lefty2* or *Pitx2* was ever observed in embryos with 3-5 somites, *Rfx3* mutants exhibit phenotypic variability in the number and length of node cilia, and overt L-R defects (Bonnafe *et al.*, 2004; Choksi *et al.*, 2014). Likewise, node expression of both *Foxj1* and *Rfx3* is greatly depleted in 8.0 dpc mouse embryos that lack the homeobox TF *Noto,* with $Noto^{-/-}$ embryos exhibiting short, malformed and immotile cilia (Beckers *et al.*, 2007). Moreover, the cilia phenotype of *Noto* null embryos can be rescued by *Foxj1* insertion into the *Noto* locus (Alten *et al.*, 2012), suggesting that NOTO acts upstream of *Foxj1* and *Rfx3*. Mutation of *Noto* does not affect A-P or D-V

patterning and its function seems specific to the L-R axis component of node activity. In combination with the compromised expression of *Noto*, *Rfx3* and *Foxj1* in *Zic2*$^{Ku/Ku}$ embryos, this suggests that ZIC2 acts upstream of these transcription factors to initiate node ciliogenesis (Figure 3.11).

At least two possible factors may account for the more severe phenotype seen in *Zic2* mutants. First, it may be that all *Zic2*$^{Ku/Ku}$ embryos have no nodal flow (whereas some *Noto*, *Foxj1* and *Rfx3* mutants generate at least a small amount of nodal flow). This could occur because of a difference in the target gene sets of these transcription factors, or because of genetic background effects. Experiments that examine the precise structure of the dysmorphic cilia or that examine cilia motility or measure nodal flow in the *Zic2*$^{Ku/Ku}$ mutants have not been conducted. Although the cilia of embryos lacking *Noto* or *Foxj1* have been shown to be short, exhibit structural defects and be mainly immotile (Alten *et al.*, 2012; Beckers *et al.*, 2007), nodal flow has not been measured in these mutants. To fully explore the hypothesis that nodal flow is more severely compromised in the *Zic2* mutants than in *Noto*, *Foxj1* or *Rfx3* mutants all alleles would need to be bred onto the same genetic background and cilia morphology, motility and nodal flow systematically compared.

A second possible reason for the difference between the phenotype of the *Zic2* mutants and that of embryos lacking *Noto*, *Foxj1* or *Rfx3* can be elucidated from similarities between the phenotype of *Zic2*$^{Ku/Ku}$ embryos and that of embryos with loss-of-function mutations of either *Pkd1l1* or *Pkd2* (Ermakov *et al.*, 2009; Field *et al.*, 2011; Pennekamp *et al.*, 2002). Loss-of-function mutations in either of these genes result in a failure of the Nodal cascade, rather than in bi-lateral or random activation of the cascade. These related proteins (PKD1L1 and PKD2) physically interact, are co-localised on cilia and are proposed to sense the left-biased signal at the node (Field *et al.*, 2011; Yoshiba *et al.*, 2012). Despite both genes being expressed in the embryonic node at a time where the L-R axis is being established, the nodes of *Pkd1l1* and *Pkd2* mutants are morphologically normal and cilia are motile, in sufficient numbers and of comparable length to wildtype littermates (Pennekamp *et al.*, 2002). The finding that *Pkd1l1* expression is depleted at the node of *Zic2*$^{Ku/Ku}$ embryos, suggests that the cilia of the *Ku* node may be compromised in their ability to respond to whatever nodal flow is generated in these embryos. It is possible that the failure of stereotypic cardiac situs in *Zic2*$^{Ku/Ku}$ embryos is caused by the cumulative effect of decreased nodal flow and decreased perception of flow, and thus positions ZIC2 upstream of all genes analysed in this chapter (Figure 3.11).

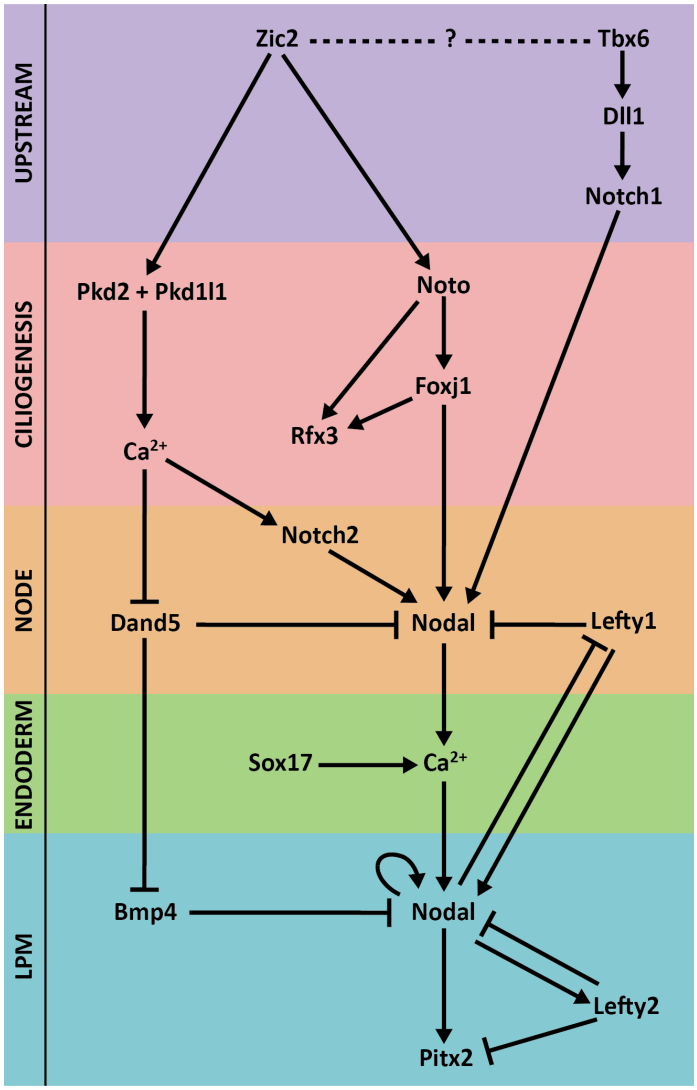


Figure 3.11: ZIC2 acts upstream of ciliogenesis and the nodal cascade to influence left-right axis formation. In the mouse embryonic node**,** NOTO, RFX3 and FOXJ1 work in synergy to form the motile and non-motile cilia required for the formation and the sensing of nodal flow. PKD2 complexes with PKD1L1 in non-motile cilia of crown cells in order to sense Ca^{2+} spikes brought about by nodal flow. This results in downregulation of *Dand5* expression and upregulation of *Nodal* expression in the left of the node. *Nodal* expression in the crown cells of the node is promoted further by Notch signalling in the primitive streak (TBX6, DLL1, NOTCH1), presomitic mesoderm (TBX6, DLL1, NOTCH1) and node (NOTCH2). Via further Ca^{2+} signalling, this leftward signal travels through gap-junctions of left endodermal cells to reach the left LPM and activate the Nodal cascade. LEFTY1 creates a barrier down the midline of the embryo, preventing signals from travelling to the right. Antagonism of NODAL by BMP4 in the right LPM further restricts the Nodal cascade to the left. Activation of the Nodal cascade in the left, but not the right of the embryo, results in organ asymmetries and the establishment of the left-right axis. I propose that ZIC2 acts upstream of ciliogenesis to influence the formation of the left-right axis, possibly via interactions with TBX6 prior to node development.

3.3.2 ZIC2 may regulate cilia retrograde transport

Though ZIC2 is required for the correct formation of cilia during embryogenesis, the ZIC2 protein itself does not appear to be part of the ciliogenesis machinery required to form functioning cilia as it was not localised in LLCPK1 cilia. ZIC2 may only localise to cilia when in the presence of a co-factor, similar to PKD2 and PKD1L1 which both show increased cilia localisation when expressed together (Kamura *et al.*, 2011). As yet, no candidates for a ZIC2 co-factor in this process have been identified. It is more likely, however, that ZIC2 acts upstream of genes involved in cilia formation. As ciliogenesis occurs in four stages (generation of centrioles, migration of duplicated centrioles, formation of basal-body structures and elongation of cilia; Hagiwara et al. 2004), the short length and bulging of *Zic2*$^{Ku/Ku}$ cilia suggests an error is occurring in the final process: cilia elongation. Short and bulging cilia appear to be common when retrograde transport is defective, with mutations in IFT144 and DYNC2H1 resulting in bulges due to accumulation of IFT particles at the ciliary tip (Liem *et al.*, 2012; Ocbina *et al.*, 2009). The bulging seen in *Zic2*$^{Ku/Ku}$ cilia is most likely 'swollen cilia', where excess cytoplasmic matrix, tubulin molecules and axoneme materials accumulate in various parts of the ciliary shaft. In normal cilia, these swollen portions tend to disappear as the cilium elongates (Hagiwara *et al.*, 2004), however, if ZIC2 acts upstream to regulate the expression of genes involved in cilia retrograde transport this may not be possible due to the shorter cilia length in the absence of ZIC2.

The creation of a Zic2-*Ku* LLCPK1 cell line via CRISPR-Cas9 mutagenesis would allow for continued analysis of the *Kumba* phenotype. It is expected that a *Kumba* cell line would replicate the cilia phenotype seen in the mouse model, and thus would provide an opportunity to investigate how components of the IFT motor protein complexes are effected by ZIC2-*Ku*. Via these cells, we can elucidate whether ciliogenesis components are completely absent, present in smaller concentrations than required, or localising incorrectly in ZIC2-*Ku* cells. The creation of a *Kumba* cell line would also allow for the analysis of whether the addition of exogenous ZIC2, or a downstream protein such as NOTO or FOXJ1, is enough to rescue the defective cilia phenotype. The mutation of *ZIC2*, or even the removal of *ZIC2*, from a ciliated cell line would allow us to determine if ZIC2 is essential for the complete formation of cilia.

As ZIC2 does not appear to localise to cilia, it's influence on cilia development most likely occur earlier in embryonic node and cilia development than the stages examined here. Therefore, future investigation should aim to identify the influence of ZIC2 on genes upstream of those already examined to pinpoint at what time point ZIC2 is essential for L/R axis establishment. The induction of *Nodal* expression in the crown cells occurs via the interactions of NOTCH1 and NOTCH2 receptors with upstream components TBX6 and DLL1 in the primitive streak and

presomitic mesoderm, with *Tbx6*$^{-/-}$ and *Dll1*$^{-/-}$ mice exhibiting multiple L-R defects (Hadjantonakis *et al.*, 2008; Krebs *et al.*, 2003). The L-R phenotype reported for *Tbx6*$^{-/-}$ embryos closely resembles that of *Zic2*$^{Ku/Ku}$ embryos, suggesting a possible interaction between ZIC2 and TBX6 prior to node development (Figure 3.11). As *Nodal* expression is aberrant in *Kumba* embryos, analysis of *Tbx6* and *Dll1* expression in *Zic2*$^{Ku/Ku}$ embryos will determine whether this defect is due to an inability to induce *Nodal* expression via Notch signalling during node and cilia development, or whether ZIC2 exerts its influence further upstream.

3.3.3 Congenital heart malformations and *situs inversus*

Congenital cardiac malformations occur in ~1% of all live births, often linked to other congenital diseases such as Heterotaxy or *situs inversus* (Gelb, 2004). The work presented here indicates that a severe loss-of-function *Zic2* allele is associated with cardiac abnormalities. The overt L-R axis defects seen in *Zic2*$^{Ku/Ku}$ embryos not only included hearts that exhibited sinistral looping, but also hearts that looped ventrally. As 60-100% of Heterotaxy patients present with atrioventricular septal defects, it is likely that a subset of the *Zic2*$^{Ku/Ku}$ embryos will also. Less common are abnormalities of systemic and pulmonary venous drainage, malpositioning of the great vessels, and subpulmonary or aortic obstructions (Brueckner, 2007; Zhu *et al.*, 2006). Thus, further investigation into the hearts of *Zic2*$^{Ku/Ku}$ embryos via histological analysis will determine the full extent of heart defects caused by *Zic2* mutation, and may reveal an underlying cause for the early demise of *Zic2*$^{Ku/Ku}$ embryos.

The reduction in the Nodal cascade and the defects seen in node cilia in *Zic2*$^{Ku/Ku}$ embryos suggests that the inability to establish a correct L-R axis is not restricted to the heart. This is further reinforced by the randomized *Pitx2* expression seen in the cardinal vein. Thus, investigation into the asymmetries of other organs should be investigated. Though *Zic2*$^{Ku/Ku}$ embryos die at 13.5 dpc, the positioning of the spleen and liver in respect to the midline can be discerned prior to this. Additional histological analysis, or imaging via micro-CT or Optical Projection Tomography (Hiraiwa *et al.*, 2013; Quintana and Sharpe, 2011), of 11-12 dpc embryos will allow for investigation into the asymmetries of the other organs. Additionally, the defective cilia in the node of *Zic2* mutants raises the possibility of defects in other ciliated organs such as the trachea or kidneys. Therefore, antibodies that mark cilia components such as α-Acetylated Tubulin or cilia-associated polycyctins can be used to determine if the defective cilia in *Zic2*$^{Ku/Ku}$ embryos are restricted to the node.

3.3.4 Are defective cilia causative for other *Zic2*-associated co-morbidities?

Spina bifida, one of the most common forms of neural tube defect (NTD) (Au *et al.*, 2010), results from failure of the posterior neuropore to completely close (Greene and Copp, 2009). The neural

tube forms along the dorsal midline of the embryo via rolling up and fusion of the neural plate (or neuroepithelium), made from thickened ectodermal cells overlying the mesoderm (Greene and Copp, 2009; Kibar *et al.*, 2007). Patterning of the dorsoventral axis of the neural tube occurs via SHH signalling (Murdoch and Copp, 2010), along with specification of neural cell fates (reviewed in Bay & Caspary 2012). Neural tube closure is initiated at ~5 somites at the future site of the cervical-hindbrain boundary and continues both rostrally and caudally along the length of the embryo. Tube formation is complete with the closure of the posterior neuropore at ~30 somites (Greene and Copp, 2009). Cilia line the ventricular zone of the neural tube and extend into the lumen (Bay and Caspary, 2012; Huangfu and Anderson, 2005). Here, the cilia are thought to transduce the SHH signalling cascade in the dorsal neural tube via the interactions of SMO, a transmembrane receptor, and the transcription factors GLI2 and GLI3. The cilia may also play a role in transduction of WNT and BMP signalling in the ventral tube (reviewed in Murdoch & Copp 2010), however the mechanisms behind signal transduction in the neural tube is still largely unknown, and it is undetermined whether neural tube cilia are motile or if they can sense secreted ligands in the neural tube lumen.

Though there is currently no evidence explicitly linking spina bifida to cilia defects, the link between cilia defects and neural tube defects via abrogated signalling pathways is strong. Conditional *Arl13b* mutants have been shown to exhibit spina bifida, most likely due to abnormal SHH signalling in the neural tube. ARL13B, a GTPase that localises to the cilia, leads to significantly shorter cilia when mutated. This in turn results in low levels of GLI activators in the neural tube compared to wildtype embryos, and thus aberrant SHH signalling (Caspary *et al.*, 2007). Whether the spina bifida in *Arl13b* mutants develops secondary to other defects such as exencephaly is still unknown, however a link between defective ciliogenesis and spina bifida is still presented. Furthermore, mutant embryos for the cilia-localised polycystin PKD1 exhibit abnormally sized renal cilia, as well as spina bifida (Lu, 2001; Nikonova *et al.*, 2014), providing further evidence for a possible link between the two phenotypes. Additionally, hydrocephaly, an established phenotype of primary ciliary dyskinesia (Norris and Grimes, 2012), frequently co-occurs with spina bifida aperta (Greene and Copp, 2009), suggesting a common underlying cause. Hydrocephaly develops when cerebral spinal fluid (CSF) accumulates within the cerebral ventricles. Motile cilia are required to move CSF through the cerebral aqueduct and ventricles, with cilia defects resulting in a build up of CSF in the brain (reviewed in Sotak & Gleeson 2012; Ibañez-Tallon et al. 2004).

Spina bifida aperta is regularly detected in murine alleles of *Zic2*, with abnormal folding of the posterior neuropore at 9.5 dpc the origin of the defect (Elms *et al.*, 2003; Nagai *et al.*, 2000; Ybot-Gonzalez *et al.*, 2007). Furthermore, $Zic1^{+/-};Zic2^{Kd/+}$ compound heterozygotes exhibit

hydrocephalus (Aruga *et al.*, 2002; Inoue *et al.*, 2004). As the underlying molecular cause of spina bifida in *Zic2* mutants is unknown, there is an opportunity to investigate whether cilia in the *Kumba* neural tube resemble those of wildtype embryos.

3.3.5 HPE and Heterotaxy

Like the *Zic2*-associated HPE phenotype, cardiac abnormalities are found only in embryos homozygous for the *Ku* allele. Human patients with two mutated copies of *ZIC2* have never been identified and HPE can clearly be caused by loss-of-function of one copy of *ZIC2* (Roessler *et al.*, 2009a). The reason for this apparent discrepancy between human and mouse genetics is not known, but it is commonly observed that mice are less sensitive to haploinsufficiency than humans (Bogani *et al.*, 2005). As 50% of patients with immotile or abnormal cilia motility exhibit *situs inversus* (Norris, 2012), it seems likely that at least a subset of human HPE patients with mutation of *ZIC2* will have cardiac and/or other defects of L-R axis formation and this is supported by the current clinical data. Whilst the exact manifestations of non-forebrain anomalies in human patients are not well documented, it is possible that a subset may be due to defects in cilia which are known to be involved in the development and function of each of the affected body systems (reviewed in Fliegauf et al. 2007; Waters and Beales 2011). Thus, ZIC2 may be an as yet unidentified influence for ciliopathies. Moreover, the observed incidence of *ZIC2*-associated cardiac defects would likely increase if this feature is specifically examined. Children presenting with *ZIC2*-associated HPE require examination for cardiac and other visceral *situs* problems. It is also possible that mutations in *ZIC2* will emerge as a risk factor for Heterotaxy once genome sequencing (rather than gene specific mutation detection) of proband DNA becomes the norm.

3.3.6 Conclusion

The asymmetries exhibited in vertebrate organ systems are derived from an initial symmetry-breaking event during early embryogenesis, whereby cilia in the embryonic node generate a leftward flow, turning on the Nodal cascade in the left, but not the right, of the embryo. Errors in any number of the components involved in establishing this L-R axis result in L-R and cardiac defects. The work in this Chapter implicates *Zic2* in cilia and L-R axis defects for the first time, providing an aetiology for the cardiac defects seen in HPE patients and a new aim for phenotyping and genetic screens of human HPE and Heterotaxy probands. As ZIC2 does not appear to directly localise to the cilium, it most likely regulates components involved in L-R determination much earlier than node development and ciliogenesis. As such, future investigations should focus on the role of ZIC2 prior to node establishment at 7.25 dpc.

Chapter 4: In silico analysis of the *ZIC2* NCE ER element

4.1 Introduction

As described in the Introduction (Section 1.11), in 2012 Roessler et al. used comparative genomics to identify a NCE (and thus a putative regulatory element) within the *ZIC2* 3'UTR. A case-control study of variant sequences across the NCE demonstrated the element is under strong selective pressure and supported the hypothesis that it regulates *ZIC2* expression in a manner that prevents HPE occurrence (Roessler *et al.*, 2012a). Gene expression can be regulated by controlling the rate of transcription, a process that is dependent upon the interaction of the gene's promoter with non-coding, *cis*-acting DNA regulatory elements, independent of their position or orientation in the genome (Blackwood and Kadonaga, 1998; Katzman *et al.*, 2007; Woolfe *et al.*, 2005). Such regulatory elements are made up of clusters of transcription factor binding sites (TFBSs) and are termed enhancers (E) or repressors (R) depending upon whether their net influence up- or down-regulates transcription, respectively (Blackwood and Kadonaga, 1998; Borok *et al.*, 2010). ER elements were first discovered when SV40 viral repeat sequences were shown to increase expression of a rabbit *β-globin* construct in HeLa cells (Banerji *et al.*, 1981). Subsequently, their ability to control gene expression in the human genome in a tissue specific manner was also demonstrated (Banerji *et al.*, 1983).

It is estimated that over 1 million ER elements exist in the human genome (Heintzman *et al.*, 2009) and that ~93% of SNVs associated with complex disease lie within non-coding regulatory elements that alter the spatial and temporal expression of their target genes in a tissue specific fashion (Maurano et al. 2012). Both enhancer and repressor activity can influence phenotype. Genetic screening of autism patients identified a SNV in an ER element that upregulates *Dlx5* and *Dlx6* expression in the di- and telencephalon. The SNV was shown to inhibit enhancer activity in transgenic murine embryos, leading to a decrease in *lacZ* reporter expression in the telencephalon, and reduced binding affinity for DLX1/2 proteins in vitro at both a site containing the SNV and a site nearby the SNV (Poitras *et al.*, 2010; Zerucha *et al.*, 2000). Similarly, an ER element in the 5'UTR of the growth-hormone receptor (*GHR*) gene was shown to regulate *GHR* expression in a tissue specific manner, via the formation of a repressor protein complex in cell based assays. In mice predisposed to diabetes, repression of *GHR* by this protein complex in the kidney protects the mouse from developing insulin-like growth factor 1 mediated insulin-dependent diabetes mellitus, whereas repression of *GHR* in the liver by the same complex mediated diabetes development (Gowri *et al.*, 2003).

4.1.1 Enhancer and repressor machinery

In eukaryotes, a typical gene possesses multiple ER elements which contribute to its overall spatial and temporal expression throughout development (Blackwood and Kadonaga, 1998; Ong and Corces, 2011). Each ER element itself can contain both enhancer and repressor domains, with the net regulatory output of the element determined by integration of these domains (Arnosti and Kulkarni, 2005). This was elegantly demonstrated in transgenic *Drosophila* embryos containing compact synthetic ER elements with different combinations of TFBS for transcriptional repressors and activators (Kulkarni and Arnosti, 2003). The resulting variation in reporter expression patterns was dependant on the type, orientation and number of each class of binding site present, with the strongest expression seen when both activator and repressor domains were activated at the same time. Thus, the separate domains within an ER element can combine differently to control spatial-temporal patterns of target gene expression.

The TF proteins bound directly to the DNA of ER elements act as a foundation for protein complexes termed enhanceosomes or repressosomes. These '-somes' include the co-activators or co-inhibitors that control interactions with the general transcriptional machinery (Carey, 1998; Merika and Thanos, 2001). By receiving signals from multiple regulatory cascades, they produce one unified output targeted to a specific promoter that results in more efficient regulation than would occur from TFs acting independently (Maston *et al.*, 2006; Panne, 2008). Repressosomes are often utilised by strong promoters to establish or maintain stable levels of gene expression. In contrast, enhanceosomes are most commonly utilised by weak enhancers to increase gene expression (Struhl, 1999).

An enhanceosome recruits RNA polymerase II and the required transcriptional machinery to the promoter of the target gene, facilitating transcription of the target (Borok *et al.*, 2010). First, activating TFs bind to the ER element and then to co-activators to create a protein complex (Merika and Thanos, 2001). Enhanceosome formation is reliant on a stereotypical arrangement of TFBSs present in an ER element to allow for cooperative recruitment of co-activators. Changing the spacing between the active binding sites, or mutating a site, often has a deleterious effect on the function of the '–some' (Courey and Jia, 2001). Second, recruitment of chromatin remodelling proteins occurs, allowing the enhanceosome to bind the promoter of a target gene and initiate transcription (Visel *et al.*, 2009). The current model for how this occurs is termed DNA looping (Figure 4.1). Upon binding of TFs to the ER element, the DNA containing the element loops around to bring the ER element-protein complex in line with its target promoter (Blackwood and Kadonaga, 1998). These looping interactions have been found to be reliant on the developmental stage of the organism (Palstra *et al.*, 2003) and also require the concerted effort of lineage specific proteins (Palstra and Grosveld, 2012). At this time, nucleosomes

flanking the ER element mask the TATA box at the transcription start site. Once enhanceosome formation is complete, however, the nucleosomes become acetylated and the chromatin is remodelled, resulting in the repositioning of the nucleosomes so the TATA box and transcription start site are accessible to TATA-binding protein (TBP), RNA polymerase II and other transcriptional machinery (Agalioti *et al.*, 2000; Panne, 2008). This looping model has been shown to occur with a murine *β-globin* ER element (HS2) via FISH TRAP analysis on livers from 14.5 dpc embryos. The HS2 enhancer region loops around to bring distal enhancers into contact with their target genes in the cell nucleus, whilst the inactive genes and chromatin between the two components loops out (Carter *et al.*, 2002; Tolhuis *et al.*, 2002). When an ER element-enhanceosome complex is pre-looped to a target promoter yet unable to actively drive gene expression, it is considered poised and can modulate target transcription in response to environmental cues. Poised enhancers are primarily linked to early embryogenetic processes such as gastrulation, neurulation and mesoderm formation, and are hypothesised to acquire an active state upon cell differentiation to a specific fate (Rada-Iglesias *et al.*, 2011). Though other models, such as DNA scanning and facilitated tracking (Zhu *et al.*, 2007), have also been hypothesised to bring an ER element and '-some' to a target promoter, there is less experimental evidence in support of these. It is most likely a combination of all three theories occurring in cells, dependant on the ER element, co-factors and target genes in question.

In contrast to enhanceosomes, a repressosome prevents the recruitment of the transcriptional machinery to a target promoter. Rather than altering the recruitment of polymerase (RNA polymerase II has been found bound to many promoters even though their target gene is in a dormant state), repressosomes appear to prevent transcriptional co-activators from interacting with the target promoter (Geanacopoulos *et al.*, 1999; Hochschild and Dove, 1998; Payankaulam *et al.*, 2010). In *E. coli,* the galactose repressor, GalR, mediated by a repressosome protein complex and the protein HU, loops around and binds to tandem operators flanking the galactose operon promoter and represses transcription. GalR repressor mutants are unable to form a loop even when their operator-binding activity is maintained, and the operon remains active when bacteria are assayed for β-galactosidase (Geanacopoulos *et al.*, 1999). It has been suggested that the limits of gene expression during embryogenesis are determined by the activities of transcriptional repressors (Mannervik *et al.*, 1999), which can be long-range or short-range depending on their location and targets. Long-range repressosomes block all ER elements from interacting with a target promoter no matter how far away in the genome they are; this is often referred to as silencing. Short-range repressosomes block the function of nearby DNA-bound activators whilst not interfering with those bound more distally (Courey and Jia, 2001).

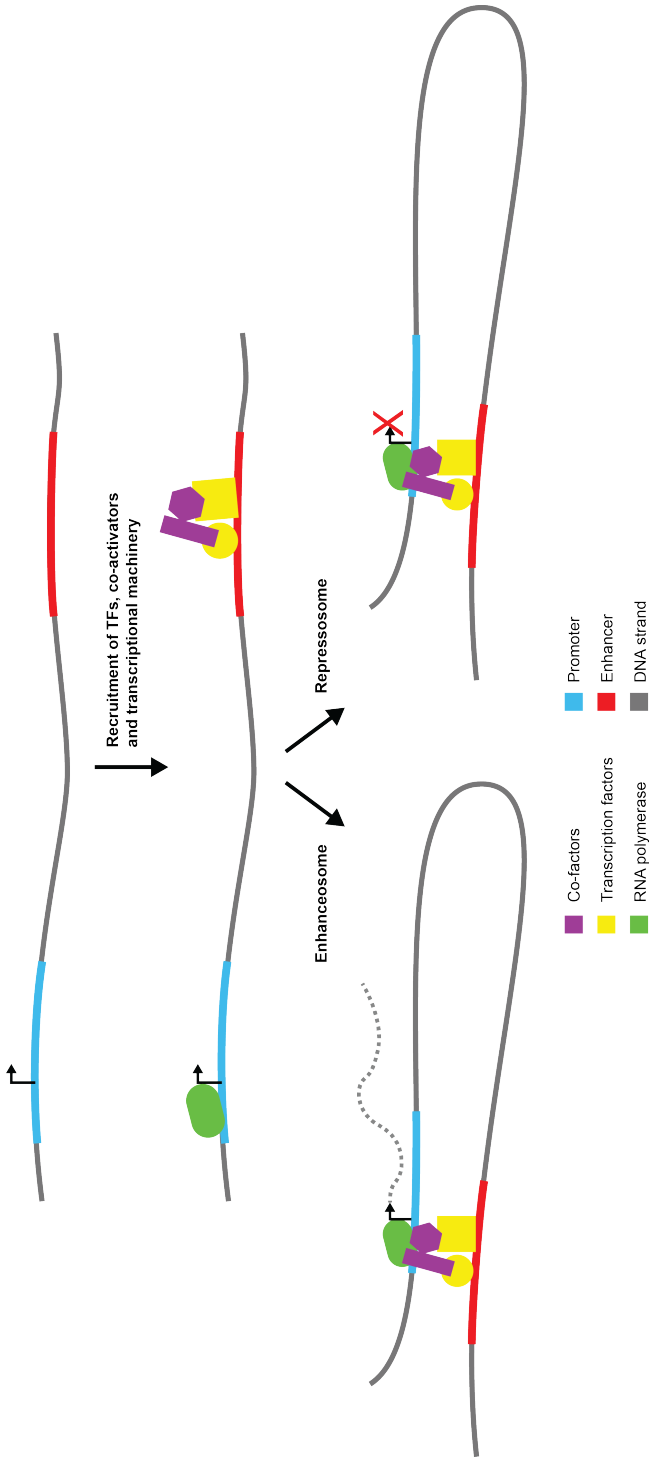


Figure 4.1: Enhanceosome and repressosome formation and function. Transcription factors and co-activators are recruited to an ER element to form either an enhanceosome complex or a repressosome complex. At the same time, chromatin remodelling proteins and transcriptional machinery (TBP, RNA polymerase II) are recruited to a target promoter. The ER element and bound protein complex loops around to interact with the target promoter and machinery. Enhanceosomes stimulate transcription of the target gene, whilst repressosomes prevent transcription.

4.1.2 The in-silico identification and analysis of ER elements

Sequence conservation is a useful tool with which to identify putative ER elements but does not provide information about whether the element is active, potential interacting TFs or the elements' chromatin state. Assessing the likelihood that a NCE may act as an ER element is therefore facilitated by combinatorial bioinformatics analyses that examine other factors.

Transcription factor binding sites

The analysis of TFBSs within a NCE can be informative for several reasons. First, clusters of TFBSs are an important indicator of an ER element as the binding of only one or a small number of TFs to a region of DNA is usually insufficient to activate or repress transcription (Shlyueva *et al.*, 2014). Second, by analysing the location and type of TFBSs it is possible to estimate how mutations in the element may alter putative DNA/TF interactions. The introduction of SNVs into an ER element may lead to the failure to form an enhanceosome and instead promote the formation of a repressosome. This was demonstrated via the disruption of binding sites in the *Ciona* Otx-a enhancer. The core 4 bp sequences of GATA and ETS binding sites were preserved, whilst the flanking nucleotides of the binding site were randomised, resulting in ectopic enhancer activity in the notochord due to the creation of new binding sites. Altering the spaces between these binding sites, or reversing the orientation of the binding sites, was also found to alter the activity of the enhancer (Farley *et al.*, 2016). Third, the particular TFs predicted to bind a region can be informative since, for example, elements required during embryo development will have a large number of binding sites for developmentally essential TFs. Genome-wide studies of DNA/protein interactions in combination with RNA-profiling experiments have repeatedly demonstrated that only a small fraction of the consensus binding sites within an ER element are typically occupied at a particular developmental time-point (Carr and Biggin, 1999; Yang *et al.*, 2006). Thus the expression patterns of the predicted TFs is also informative since TFs must be co-expressed with the target gene for functional interactions to occur (reviewed in Spitz & Furlong 2012).

Transcriptional co-activators

Transcriptional co-activators such as CREB-binding protein (CBP) and p300 simultaneously interact with the general transcriptional machinery (made up of TFIID/TFIIB, Mediator, TBP and the RNA polymerase II holoenzyme; Kuras *et al.*, 2003; Yie *et al.*, 1999), the '-somes' bound to ER elements and with the DNA of ER elements themselves to enhance transcription of a target gene. p300/CBP promote the rapid formation of the pre-initiation and re-initiation complexes, permitting multiple rounds of transcription to occur at one target gene (Yie *et al.*, 1999). Though co-activator binding sites within an ER element suggest that it is active, their absence is not

necessarily an indication of inactivity due to their interactions with the protein complexes bound to an ER element (Heintzman *et al.*, 2007; Ong and Corces, 2011; Shlyueva *et al.*, 2014).

Chromatin state

For TFs to bind an ER element, the element must be accessible to proteins. Densely positioned nucleosomes restrict TF access to ER elements, therefore nucleosome free regions are much more likely to be bound by transcriptional proteins (Shlyueva *et al.*, 2014). Measurements of chromatin state such as DNase I hypersensitivity and histone modifications (methylation and acetylation) enable prediction of an active ER region.

DNase I hypersensitivity

DNase I is an enzyme that can cleave double and single stranded DNA, including non-condensed chromatin. When chromatin is cleaved by DNase I, the underlying regions of DNA are exposed and can be accessed by a range of TFs and transcriptional machinery (Gross and Garrard, 1988). Regions of chromatin that are sensitive to cleavage by DNase I are termed DNase I hypersensitivity sites (DHSs) and can be used as markers of active *cis*-regulatory elements. DHSs have also been found to occur in promoters, silencers, and insulators, however, and alone are insufficient to indicate ER activity.

Histone modifications

Genome-wide studies that correlate histone modifications and gene expression have shown that ER elements exhibit high levels of monomethylation of histone 3 at lysine 4 (H3K4me1) and low levels of trimethylation of histone 3 at lysine 4 (H3K4me3). In contrast, promoters are usually trimethylated at this residue, so methylation type and level can be used to distinguish an active ER element from an active promoter (Heintzman *et al.*, 2007). Similarly, active ERs are marked by the acetylation of histones H3 at lysine 27 (H3K27Ac) and this has been used to distinguish between active and poised enhancers (Creyghton *et al.*, 2010; Heintzman *et al.*, 2009). A change of chromatin signature via loss of H3K27me3 and gain of H3K27Ac during cell differentiation effectively turns the enhancer on, allowing transcription of a target gene (Calo and Wysocka, 2013; Rada-Iglesias *et al.*, 2011).

4.1.3 Chapter 4 aims

The aims of this chapter were:

- To determine if the *ZIC2* 3'UTR contains the hallmarks of an active ER element via bioinformatics analyses.
- To identify candidate TFs that could bind and regulate the element at the mid-late streak node during murine gastrulation.

4.2 Results

4.2.1 Sequence conservation varies across the *ZIC2* NCE

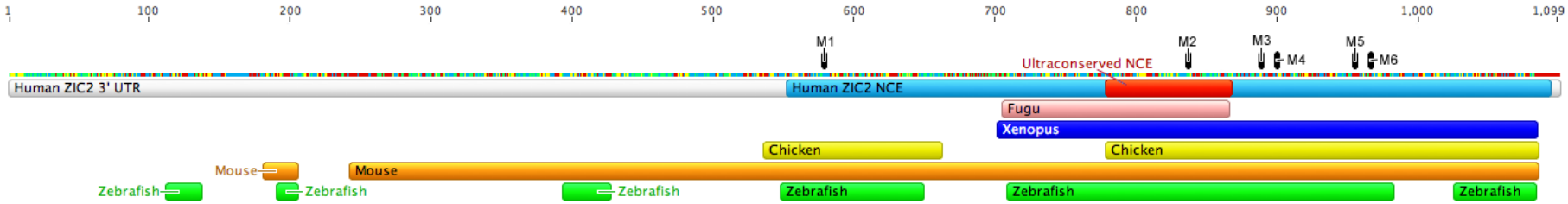
In 2012, analysis of a 10 kb region of DNA via ECRbase (Loots and Ovcharenko, 2007) and EvoPrinter (Odenwald *et al.*, 2005; Yavatkar *et al.*, 2008) identified a conserved 540 bp region in the 3'UTR of human *ZIC2*. It was hypothesised that it was an active enhancer (Roessler *et al.*, 2012a). Though Roessler *et al* investigated the sequence conservation of the region (designated NCE) in comparison to the whole 3'UTR, they did not attempt to identify further conserved sub-regions within the NCE or identify other hallmarks of an active ER element. Therefore, to further assess this region and to determine if there are ultraconserved domains within the *ZIC2* NCE, a comparison of the human *ZIC2* 3'UTR to a range of different species was performed by calculating the pairwise identity percentage between species. Whilst there is no established target percentage to determine if a sequence is evolutionarily conserved, it is agreed that the higher the pairwise identity percentage, the higher the conservation between species. Though conservation does not equate to function, the longer the conserved region is, the more likely it is to be functionally significant (Frazer *et al.*, 2003).

Sequences from multiple species were also aligned to determine whether specific regions of the NCE were more conserved than others, deemed ultraconserved regions (UCRs). Pairwise identity percentage was calculated by aligning the *Zic2* or *zic2a* 3'UTR/NCE human, mouse (*Mus musculus*), zebrafish (*Danio rerio*), *Xenopus laevis,* pufferfish (*Takifugu rubripes*), and the predicted chicken (*Gallus gallus*) sequences in CLUSTAL Omega (Goujon *et al.*, 2010; Sievers *et al.*, 2011) (Table 4.1, Figure 4.2). As expected, the mouse demonstrated the highest level of sequence conservation with human *ZIC2*. In all species examined, pairwise identity percentage in the NCE was higher than that seen in the overall 3'UTR indicating increased sequence conservation in this region of the 3'UTR. Phylogeny.fr (Dereeper *et al.*, 2008) was employed to create a maximum-likelihood phylogenetic tree for each region (Table 4.2). The phylogeny trees created in Phylogeny.fr exhibit two values: the average rate of nucleotide substitutions per site, with a longer branch representing a greater number of nucleotide substitutions from the point in which the node diverges into two sister branches, and a bootstrap value representing the percentage of times in which that node arrangement occurs in 100 computations of the tree. A value of 1.0 indicates that that node arrangement is well supported, as it occurs in every analysis of those sequences. An unrelated DNA sequence (mouse *Gapdh* CDS) was used as an anchor sequence to root the trees; it is expected that this sequence will always exist as an outgroup compared to the related sequences being analysed. The high level of NCE conservation was also reflected in the phylogenetic tree, with the human and mouse sequences consistently grouped

Table 4.1: Pairwise Identity Percentage comparison of the *ZIC2* CDS, 3'UTR and conserved regions (NCE, UCR) in multiple species. Percent identity was calculated via sequence alignment in CLUSTAL Omega 2.1 (Goujon *et al.*, 2010; Sievers *et al.*, 2011). A high identity percentage represents a high level of sequence conservation. Light green; >80% similarity. *Fugu*: *Takifugu rubripes*. CDS: coding DNA sequence, NCE: non-coding conserved element, UCR: ultraconserved region. Table corresponds to Figure 4.2.

	Accession #	CDS	3'UTR	NCE	UCR
Human *ZIC2*	NM_007129	100%	100%	100%	100%
Mouse *Zic2*	NM_009574	90.31%	82.85%	90.4%	79.27%
Xenopus zic2	NM_001087724	72.96%	56.98%	69.17%	64.94%
Zebrafish *zic2a*	NM_131558	77.75%	53.43%	59.78%	63.75%
Fugu zic2a	XM_011607711, Scaffold_31:549,069-551,451 (ENSEMBL)	74.18%	48.94%	56.36%	67.86%
Chicken *ZIC2*	XM_015274914 Predicted sequence	87.81%	NA	42.59%	46.67

Figure 4.2: Conserved domains between multiple species in the *ZIC2* 3'UTR. *Zic2* and *zic2a* 3'UTR sequences from the mouse (orange), chicken (yellow), *Xenopus* (dark blue), zebrafish (green) and *Fugu* (*Takifugu rubripes*, pink) were taken from NCBI and Ensembl and aligned to the human *ZIC2* 3'UTR (grey) in Geneious Pro (V5.5.9) to identify conserved domains. The majority of conservation was found in the NCE region (light blue), with only the mouse and zebrafish sequences aligning to the human sequence outside the NCE. One 86 bp region, containing M2, was ultraconserved (red) amongst all species. HPE-associated SNVs are annotated in black.


1
100
200
300
400
500
600
700
800
900
1,000
1,099
M1
M2
M3
M4
M5
M6
Ultraconserved NCE
Human ZIC2 3' UTR
Human ZIC2 NCE
Fugu
Xenopus
Chicken
Chicken
Mouse
Mouse
Zebrafish
Zebrafish
Zebrafish
Zebrafish
Zebrafish
Zebrafish

together. Additionally, the branches of the human and mouse sequences stemming from the nodes are relatively short, indicating fewer nucleotide substitutions took place between each sequence and a common ancestor sequence (Figure 4.3). In comparison, the chicken, pufferfish, *Xenopus* and zebrafish produced lower 3'UTR sequence conservation to the human 3'UTR. The chicken 3'UTR and NCE sequences are the least conserved to the human, and this is reflected by longer branches in the phylogenetic trees indicating a higher rate of nucleotide substitutions (Figure 4.3C). In contrast, the mouse and, interestingly, the *Xenopus* 3'UTRs and NCEs exhibit the closest relationship to the human sequences with relatively high confidence, whilst the *Fugu* and zebrafish sequences frequently form their own group. Taken together, this suggests that the NCE undergoes tighter selective constraints compared to the surrounding 3'UTR sequences, and that the relationships between species for the 3'UTR is relatively conserved.

To determine where the conserved regions were located within the 3'UTR, the sequences of each species' 3'UTR were aligned against the human *ZIC2* 3'UTR via BLASTn (NCBI) (Altschul *et al.*, 1997). Each species displayed different regions of sequence similarity when aligned to the human *ZIC2* 3'UTR, with the majority of similarity occurring within the putative NCE identified by Roessler et al (2012). One specific region of ~86 bp (human c.1599*779-864) was conserved in all species tested (Figure 4.2). This region, designated an UCR, is flanked either side by SNVs M1 and M3, with M2 occurring within the region. When analysed for pairwise identity percentage, the UCR was more conserved than the surrounding 3'UTR and CDS in the chicken, pufferfish and zebrafish (Table 4.1), suggesting a significant role for this region in evolution. Due to the small length of the UCR, a phylogenetic tree was unable to be constructed for this region.

The *ZIC3* 3'UTR was also found to contain a similar NCE (E. Roessler and M. Muenke personal communication, unpublished; Appendix Figure A2.1). Previous phylogenetic analysis indicates that the five *Zic* genes evolved from one single-copy *Zic* that underwent multiple duplication and divergent episodes early in the vertebrate lineage, and redundancies in the roles of ZIC proteins in embryo development have been documented (reviewed in Houtmeyers et al. 2013). It is therefore possible that the conserved regions in the *ZIC2* and *ZIC3* 3'UTRs stem from a common ancestor gene, resulting in similar sequences and functions. A pairwise identity percentage alignment was conducted via CLUSTAL Omega and a maximum-likelihood phylogenetic tree created via Phylogeny.fr to compare the 3'UTR sequences of all *ZIC* genes, with particular focus on the similarities in the *ZIC2* and *ZIC3* NCE.

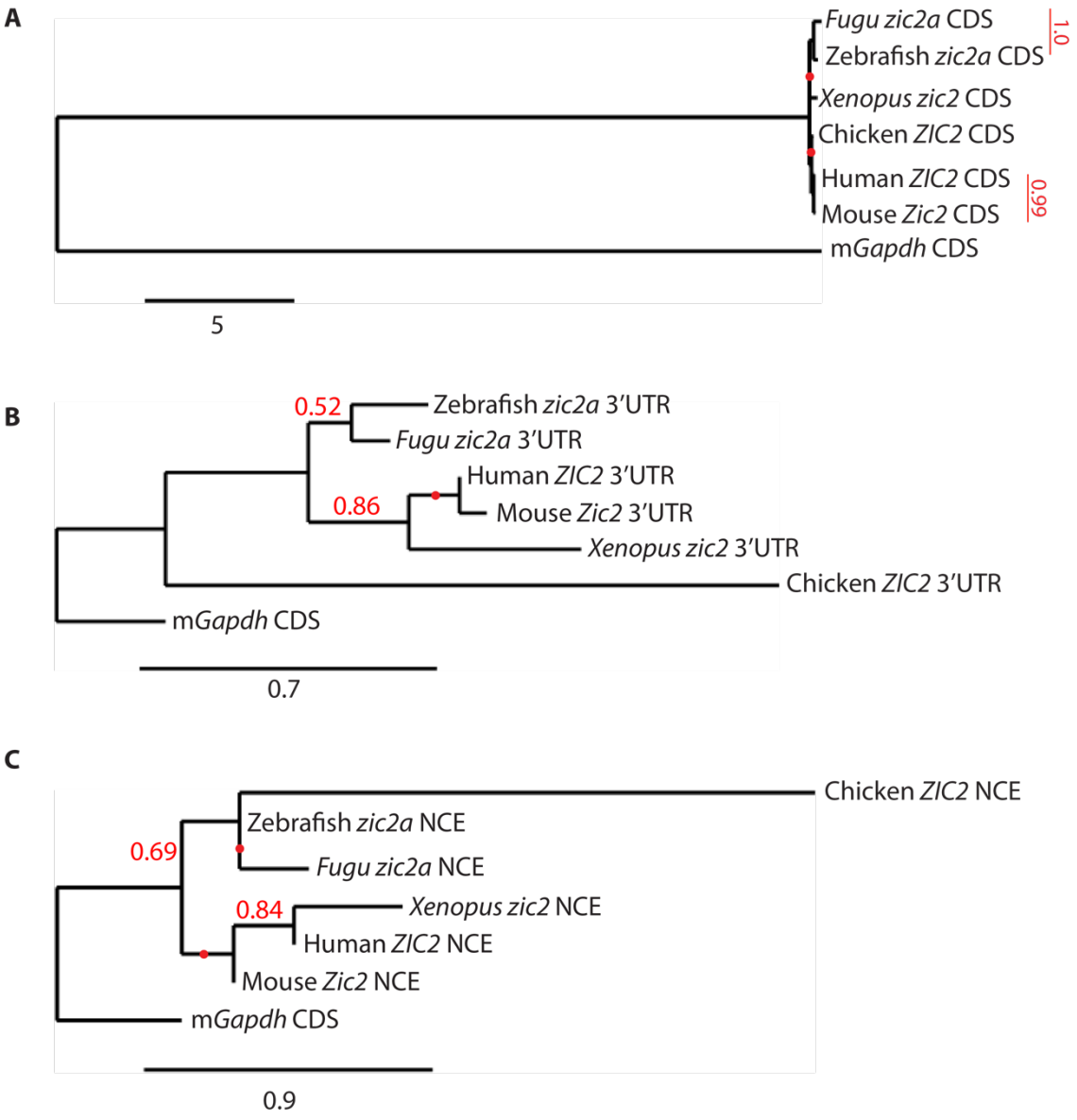


Figure 4.3: Phylogenetic comparison of *Zic2* sequences between species. Sequences were aligned in Phylogenetic.fr and a maximum-likelihood phylogenetic tree generated. **(a)** Relationship between the *Zic2* CDS, **(b)** *Zic2* 3'UTR and **(c)** *Zic2* NCE in the multiple species indicated. Scale bar represents the average rate of nucleotide substitutions per site, with a longer branch indicating a larger number of substitutions between species compared. Bootstrap values (red numbers) represent confidence in that node arrangement, with value of 1.0 signifying that that node arrangement occurs in 100% of trees computed for these sequences. Bootstrap values <50% (0.5) are indicated with a red dot and are low confidence. m*Gapdh* CDS is included as an outgroup to anchor the sequences. *Fugu*: *Takifugu rubripes.*

Both pairwise identity percentage and phylogenetic analysis indicate that the 3'UTRs of each *ZIC* gene share relatively little sequence conservation with the other *ZIC* family members, and are frequently more conserved between species than with each other (Table 4.2, Figure 4.4). For example, the human *ZIC2* 3'UTR shares 86.43% sequence conservation with the mouse *Zic2* 3'UTR, but <49% conservation with the 3'UTR of the other *ZICs*. Whilst a comparison of NCEs identified a small increase in conservation between *Zic2* and *Zic3* in both human and mouse sequences in comparison to the overall 3'UTR, the pairwise identify percentage remained relatively low (~50% for both) and both regions were pointedly more conserved between species, which is also reflected by their grouping and branch lengths of the phylogenetic trees. Overall, the low level of sequence conservation amongst the 3'UTR of the *ZIC* members and between the conserved regions of *ZIC2* and *ZIC3* suggest it is unlikely they share a similar function and they most likely evolved independently.

4.2.2 Developmental transcription factors are predicted to bind the *ZIC2* 3'UTR

As ER elements are defined by the presence of multiple TFBSs, the ENCODE genome browser (Rosenbloom et al. 2015), JASPAR (Mathelier *et al.*, 2014), and TRANSFAC databases (Matys *et al.*, 2006) were utilised to predict binding sites in the *ZIC2* 3'UTR. Though these databases can predict specific genes that bind a region, our analysis focused on gene families to take into account binding site redundancies. 465 TFBSs were predicted to occur in the *ZIC2* 3'UTR, with 267 of these occurring within the NCE region and 68 overlapping with at least one of the six HPE-associated SNVs (Table 4.3, Appendix Figure A2.2). A Chi-squared goodness of fit test found that the number of binding sites was significantly enriched in the NCE compared to the non-conserved 3'UTR region (p<0.01; Table 4.4 A), correlating with the observation that TF binding sites cluster within ER elements and are not distributed randomly within the 3'UTR. Furthermore, binding site enrichment was seen at the nucleotides for five of the six HPE-associated SNVs when compared to the remaining 3'UTR nucleotides (p<0.01; Table 4.4 B), suggesting these mutations may occur in TF binding hot spots within the NCE. The majority of predicted TFs were from developmentally important families such as SOX, FOX, CDX and HOX.

Upon introduction of the six HPE-associated SNVs into the *ZIC2* NCE, 32 binding sites were estimated to be lost (Table 4.3, Appendix Figure A2.2), with some SNVs having a greater impact than others. The introduction of M3 resulted in the loss of 14 predicted binding sites, in comparison to the introduction of M1 or M2 where only one binding site was lost for each mutation. In addition to the loss of TFBSs, the introduction of the SNVs created 18 new binding sites (Table 4.5, Appendix Figure A2.2). The creation of new binding sites, as well as the loss of binding sites, has the potential to disrupt the formation of '-somes' required for normal target

A

	m*Zic1* 3'UTR	h*ZIC1* 3'UTR	m*Zic2* 3'UTR	h*ZIC2* 3'UTR	m*Zic3* 3'UTR	h*ZIC3* 3'UTR	m*Zic4* 3'UTR	h*ZIC4* 3'UTR	m*Zic5* 3'UTR	h*ZIC5* 3'UTR
m*Zic1* 3'UTR	100.00	86.20	39.16	42.34	53.53	52.54	52.48	52.74	46.39	44.17
h*ZIC1* 3'UTR	86.20	100.00	41.70	41.09	49.60	49.93	48.34	49.27	42.11	46.38
m*Zic2* 3'UTR	39.16	41.70	100.00	86.43	43.77	43.68	48.72	48.74	NA	43.42
h*ZIC2* 3'UTR	42.34	41.09	86.43	100.00	43.09	42.76	48.08	48.74	NA	43.68
m*Zic3* 3'UTR	53.53	49.60	43.77	43.09	100.00	91.44	44.40	46.68	NA	42.32
h*ZIC3* 3'UTR	52.54	49.93	43.68	42.76	91.44	100.00	43.88	47.22	NA	41.11
m*Zic4* 3'UTR	52.48	48.34	48.72	48.08	44.40	43.88	100.00	79.53	52.67	46.91
h*ZIC4* 3'UTR	52.74	49.27	48.74	48.28	46.68	47.22	79.53	100.00	53.33	51.75
m*Zic5* 3'UTR	46.39	42.11	NA	NA	NA	NA	52.67	53.33	100.00	92.49
h*ZIC5* 3'UTR	44.17	46.38	43.42	43.68	42.32	41.11	46.91	51.75	92.49	100.00

Table 4.2: Percent Identity Matrix showing sequence similarity of annotated DNA. (a) Comparison of the human and mouse 3'UTR sequences of all *ZIC* genes. **(b)** Comparison of the human and mouse NCE sequences of *Zic2* and *Zic3*. Percent identity was calculated via sequence alignment in CLUSTAL Omega 2.1 (Goujon *et al.*, 2010; Sievers *et al.*, 2011). A high identity percentage represents a high level of sequence conservation. Green; >80% homology, yellow; 60-80% homology.

B

	m*Zic2* NCE	**h*ZIC2* NCE**	**m*Zic3* NCE**	**h*ZIC3* NCE**
m*Zic2* NCE	100.00	93.58	53.83	52.61
h*ZIC2* NCE	93.58	100.00	54.92	53.91
m*Zic3* NCE	53.83	54.92	100.00	93.65
h*ZIC3* NCE	52.61	53.91	93.65	100.00

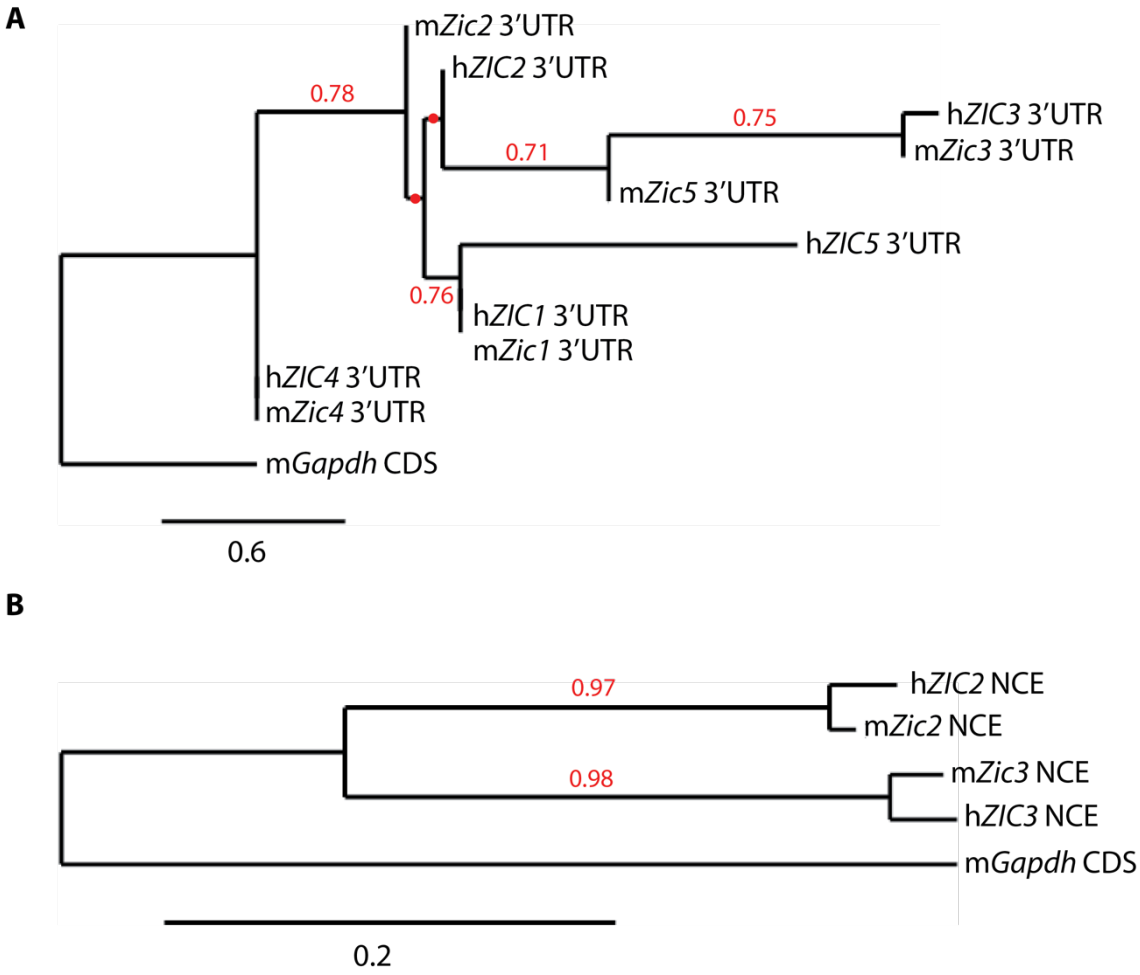


Figure 4.4: Phylogenetic comparison of the 3'UTR and NCE of *ZIC* genes. Sequences were aligned in Phylogenetic.fr and a maximum-likelihood phylogenetic tree generated. **(a)** Relationship between the mouse and human 3'UTRs of all *ZIC* family members. **(b)** Relationship between the mouse and human *ZIC2* NCE and *ZIC3* NCE. Scale bar represents the average rate of nucleotide substitutions per site, with a longer branch indicating a larger number of substitutions between species compared. Bootstrap values (red numbers) represent confidence in that node arrangement, with value of 1.0 signifying that that node arrangement occurs in 100% of trees computed for these sequences. Bootstrap values <50% (0.5) are indicated with a red dot and are low confidence. m*Gapdh* CDS is included as an outgroup to anchor the sequences.

Table 4.3: Transcription factor binding sites predicted to occur in the *ZIC2* 3'UTR, and how the introduction of SNVs alter the sites. Frequency: the number of times each transcription factor binding site is estimated to occur in the entire 3'UTR. Binding sites that are predicted to occur in the 540 bp NCE region are indicated by '+', along with any of the HPE-associated SNVs that the binding site overlaps. Lost: binding sites predicted to be lost upon the introduction of the SNVs, also indicated by '+'. Binding sites were predicted by ENCODE, JASPAR and TRANSFAC (Mathelier *et al.*, 2014; Matys *et al.*, 2006; Rosenbloom *et al.*, 2013). Table correlates to Appendix Figure A2.2.

Transcription Factor	Frequency	Present in NCE	Overlapping SNV	Lost	Transcription Factor	Frequency	Present in NCE	Overlapping SNV	Lost
AFP	1				ER	5	+		
AML	3	+			ERG	1	+		
AP	6	+			ETF	5	+	M5	
AR	7	+			ETS	2	+	M1	+
ARI	5	+	M3, M4, M5	+	EZF	1			
ARID	1	+			FOR	1	+		
ARNT	1				FOS	4	+	M3, M4	+
ARP	2				FOX	16	+	M4, M5	
BRCA	2	+			FREACT	1	+	M5	
BRN	1	+	M3		GABPA	1			
CAC	4	+			GATA	26	+	M4, M5	
CAR	1	+			GBX	3	+		
CBF	1	+	M6		GR	16	+	M3	+
CDX	5	+	M3, M4, M5	+	HES	3	+		
CEBP	5	+	M6	+	HIF1	1			
COUP	2	+			HLTF	13	+	M1	
CP	1				HMG	1			
CREB	2	+	M3, M4	+	HNF	23	+		
CRX	2	+	M3		HOX	3	+	M3, M6	+
CTCF	3				HRP	5			
DLX	1	+	M3		IL	2	+		
DUX	2	+	M3, M4	+	IPF	4	+		
E2F	3				IRF	2			
EBP	4	+			ISGF	2	+		
EGR	1				JUN	5	+	M3, M4	+
ELF	1	+			KLF	4			
ELK	3	+			LBP	5	+		
EMF	3	+	M6	+	LEF	2			
EN	1	+	M6		LF	4	+		

Transcription Factor	Frequency	Present in NCE	Overlapping Mutation	Lost	Transcription Factor	Frequency	Present in NCE	Overlapping Mutation	Lost
LHX	1	+	M3, M4	+	RFX	1	+	M5	
LUN	2				RUNX	2	+	M6	
LXR	2	+			RXR	16	+		
MAFB	1	+			SAP	2	+		
MAX	1	+			SMAD	1			
MAZ	1				SOX	11	+	M5, M6	+
MBP	1				SP	13	+		
MEF	1	+	M5		SPIB	4	+	M1	
MIT	1	+			SRF	5	+	M6	
MYB	5	+			SRY	8	+	M3, M5, M6	+
MYOD	4	+	M6	+	STAT	5	+		
MZF1	1	+			T3R	7	+		
NF	18	+	M2, M3	+	TBP	18	+	M5	+
NKX	6	+	M5, M6	+	TCF	2	+		
NOBOX	1	+	M3	+	TFIID	15	+	M5	+
PAX	4	+	M4		THAP	2	+		
PDX	1	+	M3, M4	+	TME	1	+	M5	+
PEA	2	+			TME	1	+	M5	+
PEBP	1	+			TMF	1	+		
PITX	2				TR2	1			
POU	9	+	M5	+	USF	1	+		
POU1F1	1	+			VDR	6	+		
PPUR	1				XFD	1	+		
PR	2	+			XRP	5	+		
PRRX	4	+	M3, M4, M6	+	YY	10	+	M4	
PU	1	+			ZFX	1			
PXR	1	+			ZIC	1	+		
RAR	3	+			ZNF354C	6	+		
RBP	1								

Table 4.4: Predicted transcription factor binding sites are enriched in the *ZIC2* NCE and at the HPE-associated SNVs in the wildtype *ZIC2* 3'UTR. A two-tailed Chi-squared goodness of fit test was performed to determine whether the predicted TF binding sites were randomly distributed within the 3'UTR. Significance of $p<0.05$ (*) or $p<0.01$ (**) indicates predicted binding site enrichment in the region being tested. **(a)** Binding site enrichment in the non-coding conserved element (NCE; 540 bp) compared to the non-conserved region (559 bp) of the *ZIC2* 3'UTR. **(b)** Binding site enrichment at the individual HPE-associated SNVs (1 bp) compared to the non-mutated 3'UTR (1089 bp). Observed and expected values are indicated. Degrees of freedom = 1 for all analyses.

A	Observed	Expected	χ^2	P
Non-conserved region	198	236.52	12.760	0.0004**
NCE	267	228.48		
Total:	465	465		

B	Observed (n)	Expected	χ^2	P
M1	3	0.42	15.86	0.0001**
M2	1	0.42	0.80	0.3706
M3	17	0.42	655.11	0.0001**
M4	13	0.42	377.14	0.0001**
M5	15	0.42	506.59	0.0001**
M6	12	0.42	319.57	0.0001**
Non-mutated 3'UTR	465 – mutant observed (n)	464.58		
Total:	465	465		

Table 4.5: Transcription factor binding sites predicted to be created when the six HPE-associated SNVs are introduced into the *ZIC2* 3'UTR. Table denotes the transcription factor and the respective mutation required to create the new binding site. Binding sites are predicted by ENCODE, JASPAR and TRANSFAC (Mathelier *et al.*, 2014; Matys *et al.*, 2006; Rosenbloom *et al.*, 2013).

Transcription factor	Overlapping Mutation
AR	M5
BRCA	M6
CEBP	M3
EBP	M4
FOX	M6
MYB	M6
MYC	M6
NKX	M3, M4
NP	M3
POU	M1, M3
RUNX	M4
SOX	M4
STAT	M1
TCF	M4
TFAP	M6
YY	M1
ZEB	M6

gene transcription. Not only do these new binding sites possibly interrupt the precise spacing between sites required for the formation of these complexes, but the addition or loss of a component of these complexes has the potential to change the nature of the '-some', turning an enhanceosome into a repressosome and vice versa. It is important to remember, however, that these binding sites are mere predictions, and identifying those transcription factors that bind the *ZIC2* 3'UTR requires the candidates be narrowed down further via experimental validation.

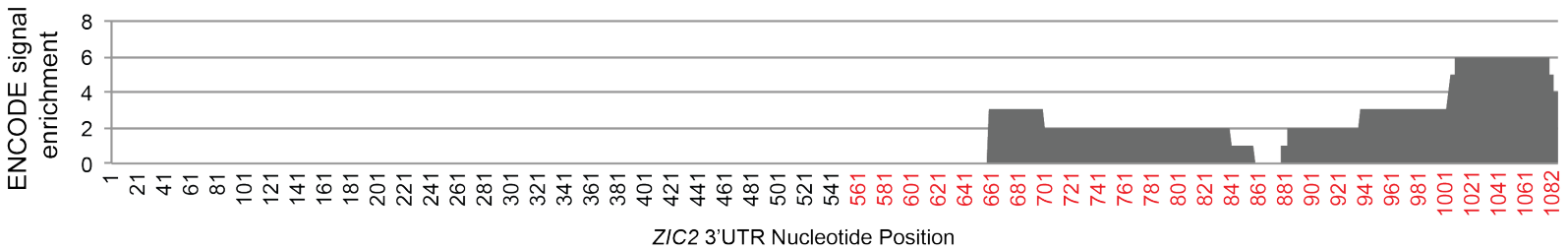
4.2.3 Co-activators are predicted to bind the *ZIC2* 3'UTR

The binding of co-activators such as p300 signals an active region of non-coding DNA. *ZIC2* 3'UTR p300 binding enrichment was analysed in GM12878 cells via the USCS Genome and Table Browser (Rosenbloom et al. 2015; SC584 Standard ChIP-seq ENCODE/SYDH). Binding was enriched in the NCE region only, with the highest peak occurring in the 3' end of the NCE (Figure 4.5 A). This peak in p300 binding coincides with a peak in monomethylation (Figure 4.5 C), a trend indicative of enhancer elements (Heintzman *et al.*, 2007). No other cells line data was available for additional analysis at this time.

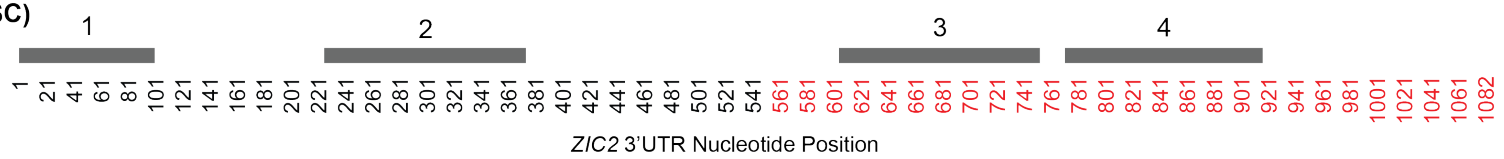
4.2.4 The *Zic2* 3'UTR chromatin state suggests it is an active element

Open chromatin allows transcriptional machinery to interact with target DNA and is indicative of an active ER element. To analyse the chromatin state of the *ZIC2* 3'UTR, the ENCODE Genome editor, which uses data from multiple microarrays (Kent *et al.*, 2002), was used to identify DHSs and the USCS Genome and Table Browser (Rosenbloom *et al.*, 2015) was used to identify histone modifications. It should be noted, however, that none of the Genome Browser annotations utilised in this analyses of the *ZIC2* NCE were conducted in the gastrulating embryo. Using ENCODE, four DHS sites were identified in the *ZIC2* 3'UTR of hESCs. As shown in Figure 4.5 B, the third and fourth DHSs overlap with the NCE, suggesting this is a region of open chromatin highly accessible to interacting proteins. Additionally, the fourth identified DHS site contains SNVs M2, M3, and M4, and overlaps with the previously identified UCR. As such, if this region is a true ER element, it is likely that M2, M3 and M4 will have a greater impact on target gene transcription than the other three SNVs based on the hypothesis that M1, M5 and M6 will be covered by a larger amount of chromatin and inaccessible to transcriptional machinery. Monomethylation (H3K4Me1), trimethylation (H3K4Me3), and acetylation (H3K27Ac) enrichment were analysed in H1-hESC in the USCS Genome and Table Browser (Rosenbloom et al. 2015). The entire *ZIC2* 3'UTR demonstrated significantly higher levels of monomethylation and acetylation compared to low levels of trimethylation (Figure 4.5 C, Figure 4.6, $p<0.05$), suggesting that the *ZIC2* 3'UTR likely contains an active ER element. Upon restricting the analysis to the

A) p300 (GM1 2878 cells)



B) DNaseI HS (hESC)



C) Histone Modifications (H1-hESC)

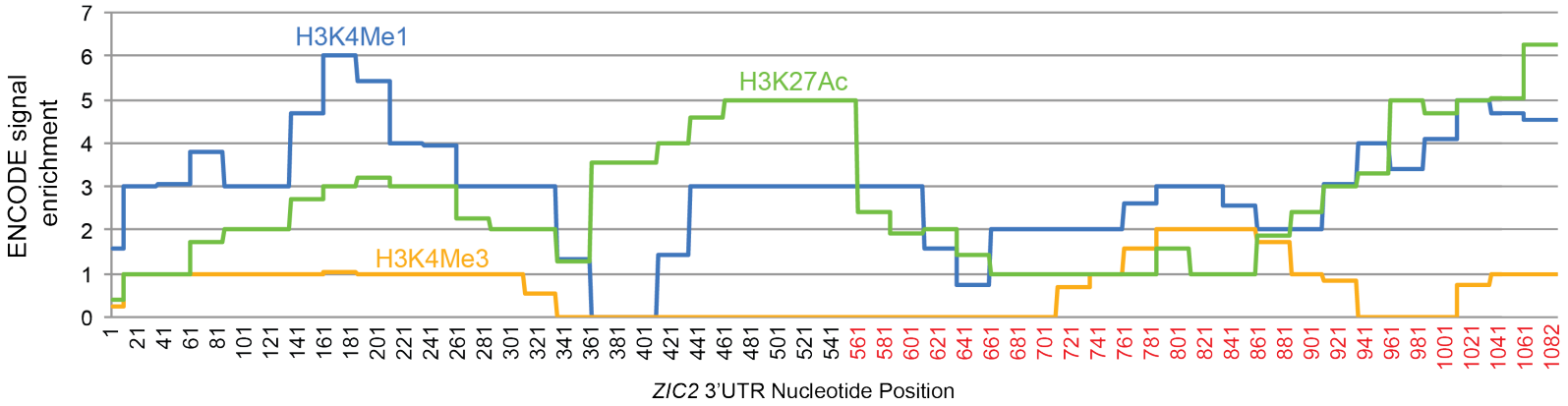


Figure 4.5: Bioinformatic analyses of chromatin state and co-activators in the *ZIC2* 3'UTR. (a) Enrichment of p300 binding sites (grey) was calculated in GM12878 cells via the USCS Genome and Table Browser (Rosenbloom et al. 2015). P300 enrichment was restricted to the NCE region. **(b)** Four DNaseI hypersensitivity sites (grey, 1-4) were detected in the *ZIC2* 3'UTR of hESCs via the ENCODE Genome browser, with two falling within the NCE (red) and one overlapping with M2, M3 and M4 (yellow). **(c)** Monomethylation (H3K4me1, blue), trimethylation (H3K4me3, orange) and acetylation (H3K27Ac, green) rates were analysed in H1-hESC using the USCS Genome and Table Browser (Rosenbloom et al. 2015). The entire 3'UTR demonstrated high levels of monomethylation and acetylation, with low levels of trimethylation. NCE nucleotide location is represented in (red).

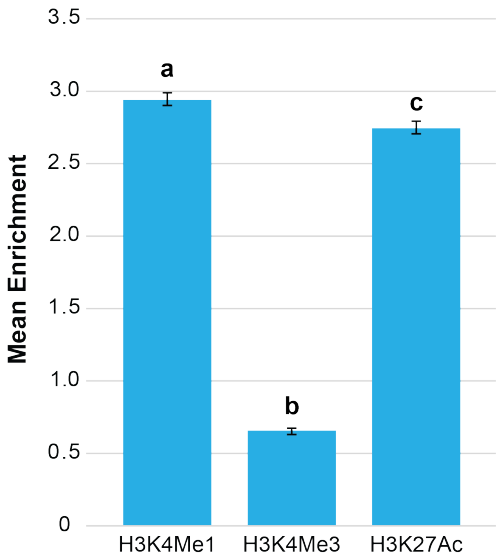


Figure 4.6: Histone modification enrichment in the *ZIC2* 3'UTR. Comparison of histone modifications in the entire *ZIC2* 3'UTR. Mean enrichment for monomethylation (H3K4me1), trimethylation (H3K4Me3), and acetylation (H3K27Ac) in the *ZIC2* 3'UTR and NCE in H1-hESC was quantified via Layered Tracks (ENCODE) in the USCS Genome and Table Browser (Rosenbloom et al. 2015). A high level of enrichment is represented as 10 and a low level of enrichment represented as 0. Enrichment within the NCE only was not significantly different from enrichment in the 3'UTR (not shown). Error bars = S.E.M. Letters denote statistical significance of $p<0.05$ calculated via a one-way ANOVA with Fischer's unprotected post ad hoc test. The letter a denotes statistical significance to b and so on.

NCE region only, or the non-conserved region of the 3'UTR, a similar result was found (not shown), suggesting the entire 3'UTR is accessible to TFs and transcriptional machinery. A track to determine if a H3K27Me3 chromatin signature was present for the *ZIC2* 3'UTR, indicating a poised state, was unavailable.

4.2.5 Candidate TFs predicted to bind the *Zic2* 3'UTR are co-expressed with *Zic2* at gastrulation

The TFs predicted to bind the *ZIC2* NCE in Section 4.2.2 are a source of interest for further studies. It is not feasible, however, to test the binding of every potential family member to the NCE. Instead, functional candidates can be narrowed down further via their spatial and temporal co-expression with *Zic2* during embryo development. As HPE develops from the defective activity of *Zic2* in the embryonic node (Warr *et al.*, 2008) gastrulation was chosen as the time point in which to analyse the expression patterns of TFs.

A combination of EMAGE gene expression databases (Richardson *et al.*, 2014) and MGI GXD databases (Finger *et al.*, 2011), which contain gene expression patterns published in peer-reviewed articles, were used to identify the expression patterns of TFs known to be expressed at gastrulation. The *Zic2* expression pattern published in Elms *et al* was used as a comparison (Elms *et al.*, 2004). Those genes that displayed overlapping or inverse expression with *Zic2* during this stage were identified as candidates to regulate the *ZIC2* NCE (Table 4.6).

The EMAGE and MGI GXD databases are made up of expression profiles from peer-reviewed articles, therefore it was not possible to analyse the expression pattern of every member of each candidate family as not all expression patterns have been published. Additionally, some candidates exhibited overlapping expression with *Zic2* at 8.5 dpc and 9.5 dpc but not at 7.25 dpc when defective *Zic2* is known to result in HPE. Whilst this suggests that these candidates may interact with the *Zic2* NCE to regulate gene expression during organogenesis, they were not considered substantial candidates for influencing HPE development at gastrulation.

Table 4.6: Proteins/genes that exhibit overlapping expression with *Zic2* during murine gastrulation. A '+' represents positive expression in that embryonic region. Most expression patterns were identified via EMAGE and MGI GXD databases (Finger *et al.*, 2011; Richardson *et al.*, 2014). Orange: enzyme; green: receptor; blue: TF; purple: secreted molecule; and red: mRNA binding protein.

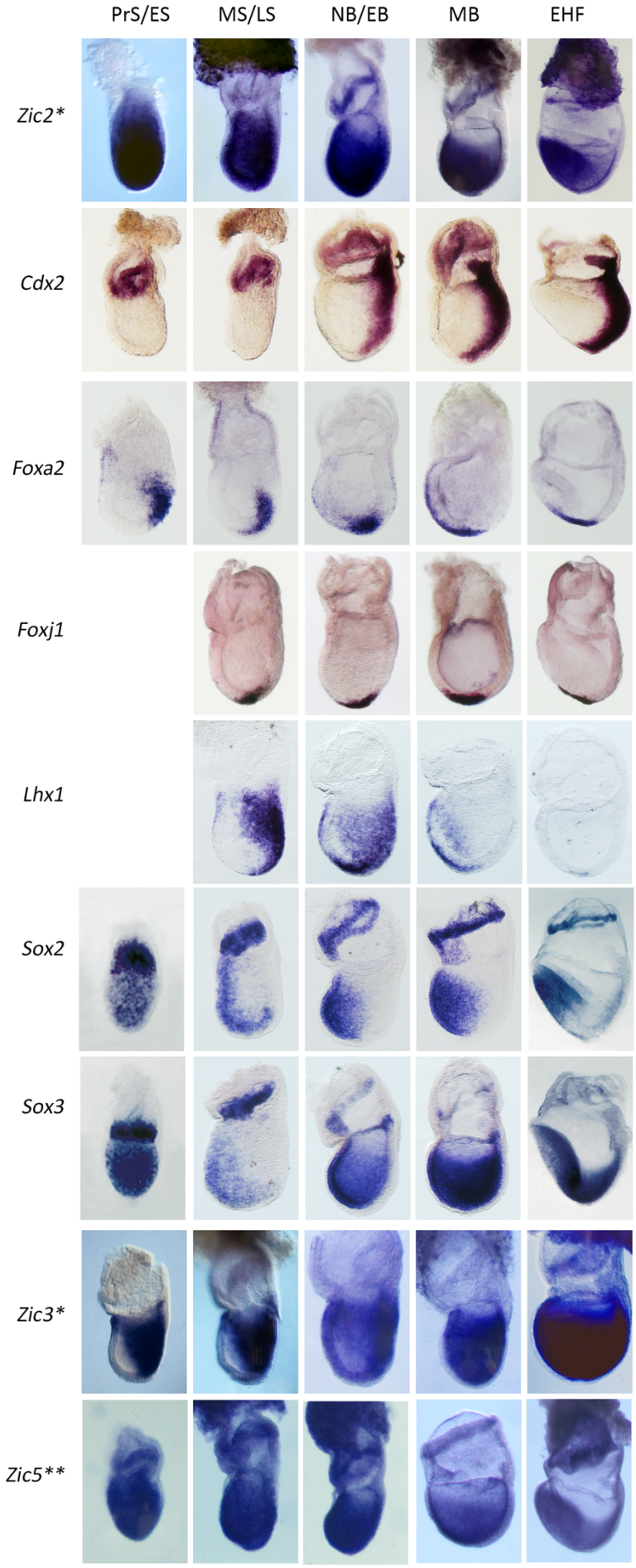
	Ectoderm	Endoderm (Definitive)	Head Folds	Head Process	Mesoderm	Neuroectoderm	Neural Groove	Neural Plate	Node	Primitive Streak	Reference
Zic2	+	+	+	+	+	+	+	+	+	+	(Elms *et al.*, 2004)
ATP9A									+		EMAGE/MGI GXD
Bicc1									+		EMAGE/MGI GXD
Bmp7					+				+	+	EMAGE/MGI GXD
CAPSL									+		EMAGE/MGI GXD
CAR3									+	+	EMAGE/MGI GXD
CELSR1									+	+	EMAGE/MGI GXD
CERL1		+			+	+					(Barratt *et al.*, 2014)
CERL2		+									EMAGE/MGI GXD
CDX1								+		+	EMAGE/MGI GXD
CDX2										+	Arkell lab (Figure 4.7)
CFC1					+				+		EMAGE/MGI GXD
CHRD									+		EMAGE/MGI GXD
CYB561									+		EMAGE/MGI GXD
DKK1		+			+	+					EMAGE/MGI GXD
DMGDH									+		EMAGE/MGI GXD
EOMES									+	+	EMAGE/MGI GXD
EPHA2									+	+	EMAGE/MGI GXD
FAM183B									+		EMAGE/MGI GXD
FGF8										+	EMAGE/MGI GXD
FGFR1	+				+	+	+	+		+	EMAGE/MGI GXD
FOXA1	+	+	+		+		+	+		+	EMAGE/MGI GXD
FOXA2				+		+			+		Arkell lab (Figure 4.7)
FOXC1					+					+	EMAGE/MGI GXD
FOXC2								+			EMAGE/MGI GXD
FOXD3						+					EMAGE/MGI GXD
FOXD4						+			+		EMAGE/MGI GXD
FOXJ1					+				+		(Barratt *et al.*, 2014)
FURIN		+							+		EMAGE/MGI GXD
GAL									+	+	EMAGE/MGI GXD
GOOSECOID									+	+	EMAGE/MGI GXD
HDC									+		EMAGE/MGI GXD
HESX1						+					EMAGE/MGI GXD
HHEX									+		EMAGE/MGI GXD
HOXB1					+				+	+	EMAGE/MGI GXD
HOXB8										+	EMAGE/MGI GXD
IFT88									+		EMAGE/MGI GXD

	Ectoderm	Endoderm (Definitive)	Head Folds	Head Process	Mesoderm	Neuroectoderm	Neural Groove	Neural Plate	Node	Primitive Streak	Reference
IGFBP5					+				+		EMAGE/MGI GXD
JOSD2									+		EMAGE/MGI GXD
LHX1		+			+				+	+	Arkell lab (Figure 4.7)
LYPD6B									+		EMAGE/MGI GXD
MESP1					+					+	EMAGE/MGI GXD
MIXL1										+	EMAGE/MGI GXD
MLF1									+		EMAGE/MGI GXD
MMP15									+		EMAGE/MGI GXD
MNX1									+		EMAGE/MGI GXD
NODAL									+		(Barratt *et al.*, 2014)
NOGGIN					+				+		EMAGE/MGI GXD
NOTCH1										+	EMAGE/MGI GXD
NOTCH2									+		EMAGE/MGI GXD
NOTO									+		(Barratt *et al.*, 2014)
OTX2		+									EMAGE/MGI GXD
PKD1L1									+		EMAGE/MGI GXD
PIM1									+	+	EMAGE/MGI GXD
PIFO									+		EMAGE/MGI GXD
PITX2					+						(Barratt *et al.*, 2014)
PLET1									+		EMAGE/MGI GXD
POPDC2									+		EMAGE/MGI GXD
PPP1R1A			+						+		EMAGE/MGI GXD
PRICKLE2									+		EMAGE/MGI GXD
PRNP			+						+		EMAGE/MGI GXD
PROX1									+		EMAGE/MGI GXD
PRRX1	+				+						EMAGE/MGI GXD
RFX3									+		(Barratt *et al.*, 2014)
SCARA3									+		EMAGE/MGI GXD
SHH	+								+		EMAGE/MGI GXD
SMIM22									+		EMAGE/MGI GXD
SMOC1									+		EMAGE/MGI GXD
SOX2	+					+					(Wood and Episkopou, 1999)
SOX3	+				+	+					(Wood and Episkopou, 1999)
SOX17		+									EMAGE/MGI GXD
T				+						+	EMAGE/MGI GXD
TGM2									+		EMAGE/MGI GXD

	Ectoderm	Endoderm (Definitive)	Head Folds	Head Process	Mesoderm	Neuroectoderm	Neural Groove	Neural Plate	Node	Primitive Streak	Reference
TRH		+	+								EMAGE/MGI GXD
TMEM176A									+	+	EMAGE/MGI GXD
VTN									+		EMAGE/MGI GXD
WNT3A										+	EMAGE/MGI GXD
WNT11									+		EMAGE/MGI GXD
ZEB1			+						+		EMAGE/MGI GXD
ZEB2			+							+	EMAGE/MGI GXD
ZIC3	+	+			+				+	+	(Elms *et al.*, 2004)
ZIC5	+		+		+						(Inoue *et al.*, 2004)

A total of 79 different genes were identified that produce overlapping or inverse expression with *Zic2* during gastrulation (Table 4.6). These candidates were further broken down by their roles in the cell (TFs, secreted molecules, receptors, enzymes or RNA binding proteins). By cross referencing these candidates with the binding sites predicted to occur in the *ZIC2* NCE (Section 4.2.2), or the binding sites of their downstream targets in the case of receptors and secreted molecules, a list of 19 genes were identified as the most likely candidates to bind and regulate the *ZIC2* NCE during gastrulation: *Cdx1, Cdx2, Foxa1, Foxa2, Foxc1, Foxc2, Foxd3, Foxd4, Foxj1, Hoxb1, Hoxb8, Lhx1, Prrx1, Rfx3, Sox2, Sox3, Sox17, Zic3 and Zic5.* As some of the identified candidates were easily obtainable or were already in pre-existing constructs in the Arkell lab, *Cdx2, Foxa2*, *Lhx1, Sox2, Sox3, Zic3* and *Zic5* were chosen for the initial analysis. *Zic2* was also considered a candidate to self-regulate its own expression. WMISH on wildtype embryos was used to confirm overlapping expression with *Zic2* during gastrulation (Figure 4.7) and candidature for the regulation of *Zic2* via NCE interactions.

Figure 4.7: The expression pattern of candidate transcription factors during mouse gastrulation. The expression patterns of *Cdx2, Foxa2, Foxj1, Lhx1, Sox2, Sox3, Zic3* and *Zic5* were analysed via WMISH and compared spatially and temporally to *Zic2* expression. In all pictures the anterior is to the left and embryos are shown in lateral view. PrS/ES: pre-streak or early streak, 5.5-6.5 dpc; MS: mid-streak, 6.75 dpc; NB/EB: no bud or early bud, 7.0-7.25 dpc; MB: mid-bud, 7.25-7.5; EHF: early head fold, 7.75. * indicates *in situs* performed by Ruth Arkell and published in (Elms *et al.*, 2004). ** indicates *in situs* performed by Ruth Arkell, unpublished. PrS/ES and EHF *Sox2* and *Sox3* in situs are adapted from (Wood and Episkopou, 1999).


PrS/ES
MS/LS
NB/EB
MB
EHF
Zic2*
Cdx2
Foxa2
Foxj1
Lhx1
Sox2
Sox3
Zic3*
Zic5**

4.3 Discussion

This Chapter extends the research by Roesseler et. al. (2012a), providing bioinformatic evidence for the *ZIC2* NCE as a putative ER element during embryonic development and identifying a UCR within the NCE. Additionally, nine candidate transcription factors likely to regulate *Zic2* expression via interactions with the NCE during gastrulation were identified.

4.3.1 The *ZIC2* 3'UTR contains hallmarks of an active ER element

Analysis of predicted DHSs sites, methylation patterns, co-factor binding and TF binding sites, along with high levels of sequence conservation amongst species, indicates that the *ZIC2* NCE is likely an ER element. Additionally, the finding of high levels of monomethylation and acetylation along with low levels of trimethylation in the *ZIC2* 3'UTR of hESCs suggest that the element is active.

The identification of an UCR within the *ZIC2* NCE provides evidence for a minimal region required for target gene regulation. If the *ZIC2* NCE is confirmed to be an enhancer or repressor via experimental methods, it will be interesting to investigate whether the UCR region can maintain this activity on its own or whether the entire NCE is required for element function. Additionally, the positioning of M2 within the UCR suggests that this mutation may produce the largest impact on element function when compared to other mutations outside of this region.

4.3.2 Candidate NCE-interacting transcription factors

Via bioinformatic analysis and gene expression profiles, the list of transcription factors predicted to bind the *ZIC2* 3'UTR was narrowed down to nine candidates, including ZIC2. All nine of the identified candidates are known to be required for correct embryo development, with a particular focus on gastrulation. Additionally, four of the nine candidates' binding sites are predicted to be lost when SNVs are introduced (CDX2, SOX2, SOX3 and LHX1), whilst four binding sites are predicted to be created upon SNV introduction (FOXA2, FOXJ1, SOX2 and SOX3). This suggests that SNVs have the potential to markedly disrupt *ZIC2* NCE activity by changing the nature of enhanceosome or repressosome formation. Further in vitro studies are required to determine if these TF candidates do indeed interact with the *ZIC2* NCE to regulate transcription, and how the SNVs affect these interactions (see Chapter 5).

One candidate TF, CDX2, is known to play an important role in somitogenesis, endoderm development and neural tube closure during embryonic development (Savory *et al.*, 2011), and is required at gastrulation for AP patterning and tissue extension via the regulation of *Hox* genes and Wnt signalling (Young *et al.*, 2009). This regulation of *Hox* genes occurs via the binding of CDX2 to two *Hoxc8* enhancers, a gene critical for early development of the neural tube and mesoderm in mice (Taylor *et al.*, 1997), however this interaction is cell specific. Mutations of the

two enhancers prevented CDX2 from binding specifically to both sites and resulted in altered transgene expression in the neural tube and somites of developing embryos (Taylor *et al.*, 1997). Similarly, the ability to bind an enhancer element has been shown for LHX1 (Costello *et al.*, 2015), FOXA2 (Gao *et al.*, 2008) and SOX2 (Yuan *et al.*, 1995). Moreover, LHX1 and FOXA2 have been shown to complex, together with OTX2 and LDB1, to regulate target gene expression and direct anterior mesendoderm, node and midline development (Costello *et al.*, 2015). Together, these studies provide precedent for the identified candidate TFs to actively bind ER elements such as the putative *ZIC2* NCE.

Whilst candidate transcription factors were analysed for their expression patterns at gastrulation and the occurrence of binding sites in the *ZIC2* 3'UTR, the list of factors identified in this chapter is far from exhaustive. Information about the expression pattern at gastrulation of a majority of the TFs identified in the bioinformatics screen could not be found. Likewise, a majority of the 79 identified TFs with overlapping gastrulation expression patterns have not had any known binding motifs published and therefore their *ZIC2* 3'UTR binding potential is unknown. Whilst this study focused on genes that met both bioinformatic and expression criteria, it does not mean that identification of candidates should be restricted to the genes listed in Section 4.2.4. Analysis of enhancers based on predicted TFBSs alone is often dependant on knowledge of the preferred binding site of the candidate proteins. The binding preferences of many candidates are unknown at this stage and the difficulty of predicting candidates is further compounded by the fact that many TF require a co-factor to bind DNA (reviewed in Andersson 2014). Furthermore, whilst many predicted TFBSs have been shown to bind their target DNA in vitro, these results often do not translate to in vivo studies as they do not take into account variables such as cell type specificity and chromatin accessibility (Andersson, 2014).

Similarly, it is important to note when analyzing binding site that short motifs frequently match to genomic or even random DNA sequences (for example, each 6 bp long motif would be expected to occur every 4^6 bp = 4,096 bp), and only a small proportion of all matches in a genome are typically bound by the corresponding transcription factor in vivo. Of these, only a small number result in regulation of gene expression (Shlyueva *et al.*, 2014; Yáñez-Cuna *et al.*, 2012). Thus, a large number of the candidate binding sites identified in this chapter are likely false-positives. Corresponding binding site predictions to gene expression patterns will remove a large number of these false positives, however there remains a possibilitiy that the final nine candidate transcription factors may not interact with the *ZIC2* NCE.

Together, this suggests that there remain multiple candidates yet to be identified. Thus, whilst it is important to investigate the role of the nine identified candidates in regulating the *ZIC2* NCE in in vitro and in vivo studies, it is also important to continue to identify new candidate genes.

Due to the nature of *cis*-acting NCEs, it is likely that the TFs that do bind the *ZIC2* 3'UTR are located nearby in the genome and that they serve a developmentally important role. As more information is acquired on gene expression patterns and binding sites are further clarified, the identification of new candidate TFs will become easier.

At this stage, FOXA2 is the only final candidate implicated in HPE development (Houtmeyers *et al.*, 2016). This affords an opportunity to identify a new selection of HPE related genes, as it is likely that the introduction of the six SNVs will alter the interactions between the final nine candidates and the *ZIC2* NCE, resulting in congenital malformations.

4.3.3 Conclusion

The identification of mutations within a putative NCE in *ZIC2* 3'UTR in HPE probands by Roessler et al (2012a) lead to the hypothesis that this element acts as an enhancer during gastrulation to regulate *ZIC2* transcription. In this Chapter I have shown that the *ZIC2* NCE contains the hallmarks of an active ER element, and have identified nine TF candidates to bind and regulate this element during gastrulation. Furthermore, I have shown that each of the six HPE-associated SNVs have the potential to disrupt the protein complexes that interact with the NCE. Though NCE function remains to be determined experimentally, it is likely an important component in *ZIC2* regulation during gastrulation and the prevention of HPE.

Chapter 5: In vitro analysis of the ER element activity of the *ZIC2* NCE

5.1 Introduction

It has recently been found that enhancers can not only switch between active, poised and repressed states, but can possess these properties simultaneously (Palstra and Grosveld, 2012). This was shown via chromosome conformation capture and ChIP on the mouse *α-globin* locus in erythroid cells and progenitors, where multiple domains within an ER element establish a poised state independently of each other through binding of specific multiprotein complexes. Additionally, multiple domains within one ER element can establish a hierarchy of activity, with the net activity of the element the determinate for gene expression. For example, an ER element situated ~10 kb upstream of the murine *Albumin* (*Alb1*) gene possess three active domains. The first acts as a positive regulator of albumin expression in rat CWSV1 cells, whilst the second acts as a repressor to negate the activity of the first. Finally, a third domain which overrides the repressor activity of the second domain was also identified (Herbst *et al.*, 1989). Moreover, transcriptional machinery complexes form at the promoter whether or not an ER element is active, and act as a docking site for the '-somes' bound to poised ER elements (Vernimmen *et al.*, 2007). By both the promoter and ER element existing in a poised state, the timing of gene expression can be delayed until the precise moment it is required in a cell, circumventing inappropriate expression that could disrupt the cell processes.

ER elements can also transition between states via changes in the TFs that interact with them. The availability of each TF is subject to the signalling pathways that are active in the same cell as the ER element. Here, they act as a 'switch', turning on transcription of a target gene in response to signalling cues. As the TFs behind these 'switches' and their target binding sites are conserved within each major signalling pathway, the 'switch' can be induced experimentally in a desired cell line via ectopic activation of the signalling pathway (Barolo and Posakony, 2002). For this to occur, developmental signalling pathways must adhere to three fundamental principles, as outlined by Barolo and Posakony (2002). First, signalling pathways exhibit activator insufficiency, whereby the binding of a single signal-modified TF to an ER element is not sufficient to activate full target gene expression. This prevents a pathway from activating all of its targets in one signalling event, reducing promiscuous activation. Instead, cooperative binding of signal-independent local activators and signal-activated factors must occur for signal-activated elements to respond, which is the second principle (Barolo and Posakony, 2002). This integration with different local activators that generates different and distinct expression patterns of a target gene, as can be seen with the NOTCH responsive protein Suppressor of

Hairless, Su(H), in *Drosophila*. By pairing with different local activators, Su(H) can bind to a variety of NOTCH responsive elements and induce expression in unrelated tissues such as the wing margin, mesectoderm cells and the socket and shaft cells of the mechanosensory bristle (Barolo *et al.*, 2000; Halder *et al.*, 1998; Morel *et al.*, 2001). By employing the first two principles, a signalling pathway specifies distinct cell fates via activation of different and overlapping subsets of target genes. Finally, signalling pathways employ default repression in the absence of signalling, ensuring target expression remains in the off state until sufficient signal is generated. This is primarily achieved by active repression of promoters or enhancers of signal-activated target genes in any conditions deemed inappropriate (Barolo and Posakony, 2002). Default repression restricts signalling to specific tissues or zones so that activation of a signal in one cell is insufficient to cause activation of all of its targets in other cells. In the case of WNT signalling, the TCF/LEF family of TFs (TCF/LEF) can mediate default repression in the absence of WNT signalling and then act as activators in signal-responding cells (Cadigan, 2008). Additionally, a fourth principle has been described in which some signalling pathways repress a gene that was active before the signal was generated (Affolter *et al.*, 2008). This directly conflicts with the third principle, as default repression would need to be overruled for a signal-dependent target gene to be active prior to signalling. Nevertheless, all the principles described contribute to the spatial-temporal regulation of gene expression that is required for correct patterning events to occur. Therefore, to establish if and how a putative ER element regulates gene expression, the element must be assessed in the correct signalling context.

5.1.1 WNT, BMP and NODAL drive anterior primitive streak and node activity at gastrulation

At the onset of gastrulation (6.0 dpc), the proximal-distal axis of the embryonic cylinder is converted into an anterior-posterior axis by the co-ordinated activities of WNT, BMP, NODAL (Figure 5.1) and FGF signalling pathways (not shown)(reviewed in Tam et al. 2006 and Pfister et al. 2007). In general, the positive (ligand, agonist) and negative (antagonist) components of each pathway are expressed at opposite sides of the embryo to create a morphogen gradient, with each cell within a tissue subject to a specific morphogen dose. *Wnt3*, for example, is expressed in the posterior PS throughout gastrulation whilst it's antagonist, *Dkk1*, is expressed in the AVE, creating a gradient of WNT signalling activity across the embryo (Ben-Haim *et al.*, 2006; Kemp *et al.*, 2005; Pfister *et al.*, 2007) that can be visualised in vivo with a transgenic reporter such as that in the TCF/Lef:H2B-GFP mouse strain (Ferrer-Vaquer *et al.*, 2010).

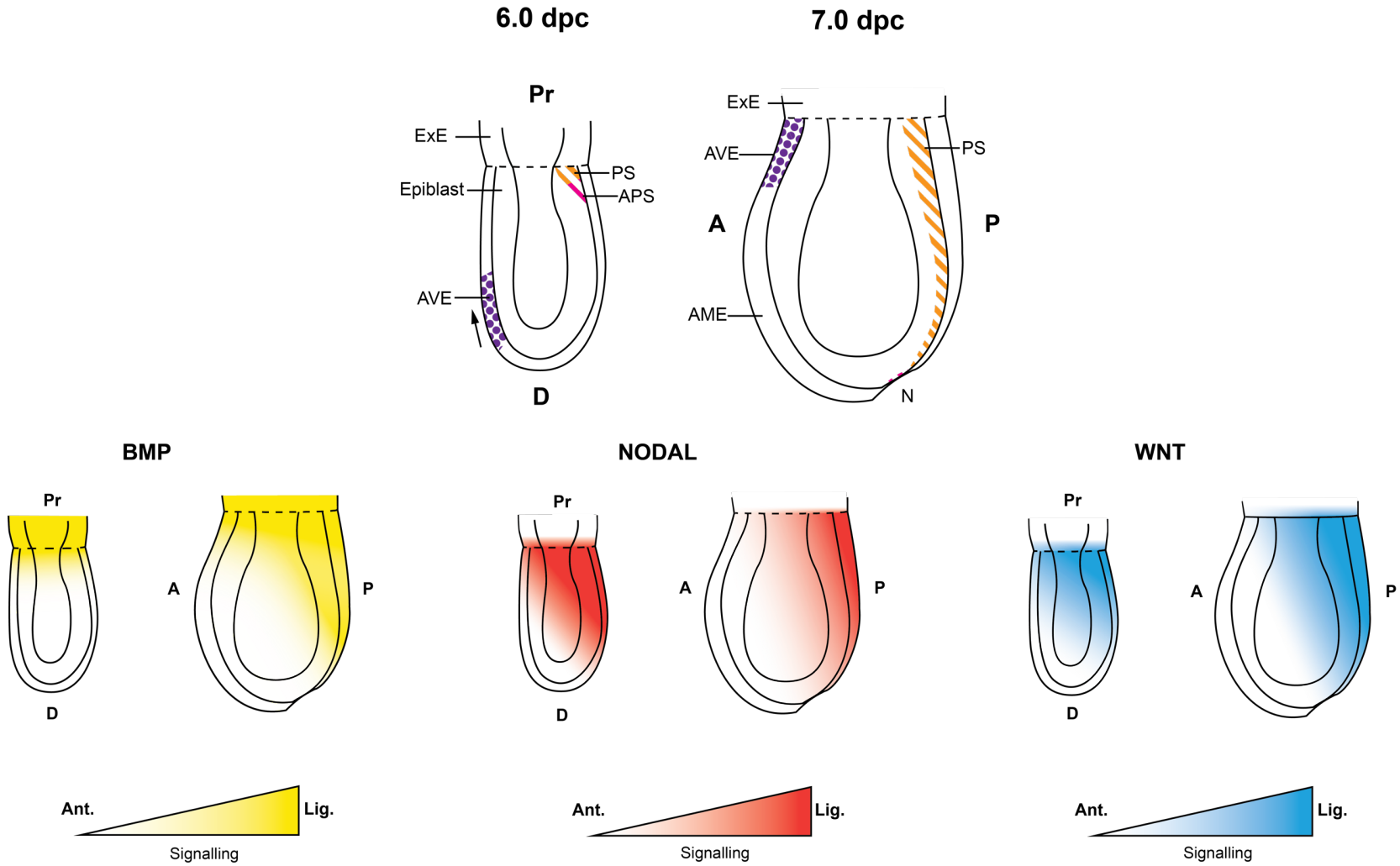

6.0 dpc
7.0 dpc
Pr
ExE
Epiblast
PS
APS
AVE
D
ExE
AVE
PS
A
P
AME
N
BMP
NODAL
WNT
Pr
D
A
P
Ant.
Lig.
Signalling

Figure 5.1: Signalling pathway activity at early-mid gastrulation. Gastrulation is initiated (6.0 dpc) in the mouse embryo when BMP4 signalling from the extraembryonic ectoderm (ExE) and visceral endoderm (VE), in conjunction with NODAL and WNT3 signalling from the ExE, VE and proximal epiblast, induces primordial germ cells to move distally to the posterior epiblast and prospective posterior of the embryo (reviewed in Tam et al. 2006; Pfister et al. 2007). BMP4 and WNT3 activity in the posterior VE activates *T* expression, thereby promoting formation of the primitive streak (Ben-Haim *et al.*, 2006; Pfister *et al.*, 2007). The AVE suppress a posterior fate in the underlying anterior ectoderm via expression of TGFβ antagonists such as *Cer1* and *Lefty1* and WNT antagonists such as *Dkk1,* causing the gradient of signalling activity to adopt an A-P orientation (Kemp *et al.*, 2005; Perea-Gomez *et al.*, 2002; Robb and Tam, 2004). This combination of a high WNT and high NODAL environment in the posterior epiblast, mediated by BMP signalling, that correctly positions the PS and induces mesoderm formation (Behringer *et al.*, 1999; Ben-Haim *et al.*, 2006; Rivera-Pérez and Magnuson, 2005). The mesoderm expands outwards from both sides of the PS and distally towards the tip of the egg cylinder and future node to form mesodermal 'wings' (Tam and Behringer, 1997) via continuous recruitment of epiblast cells (reviewed in Rivera-Perez and Hadjantonakis 2014). The APS is composed of transitory populations of tissue precursors that will give rise to the axial mesendoderm that will form the PrCP, notochord and node (Robb and Tam, 2004) and the ADE (Lawson *et al.*, 1991; Tam and Beddington, 1992; Wells and Melton, 1999), as described in Chapter 1. Formation of the ADE is favoured by high levels of NODAL signalling in the APS (Kubo, 2004; Lewis and Tam, 2006; Norris *et al.*, 2002). In contrast, formation and maintenance of the node at the mid-late streak stage (6.75-7.0 dpc) requires medium-high levels of NODAL signalling in conjunction with low levels of both WNT and BMP activity (reviewed in Davidson & Tam 2000). Downstream NODAL targets such as *Gsc* and *Foxa2* are also expressed in the node tissue, where they are required for the generation of the left-right axis, PrCP and notochord, respectively (Ang and Rossant, 1994; Beddington, 1994; Brennan *et al.*, 2002; Yamanaka *et al.*, 2007). As cells move out from the node to form the notochord, cells from the APS are recruited to repopulate the node (Kinder *et al.*, 2001). ExE: extraembryonic ectoderm, AVE: anterior visceral endoderm, PS: primitive streak, APS: anterior primitive streak, AVE: anterior visceral endoderm, AME: anterior mesoendoderm, N: node, Pr: proximal, D: distal, A: anterior, P: posterior, Ant: antagonists, Lig: ligands.

Currently, no reliable reporter strains exist for BMP or NODAL signalling, therefore their morphogen gradient must be inferred from gastrula expression analysis and mutational screens. The morphogen dose a cell receives indicates its position within the embryo and directs future differentiation, movement and morphogenesis. Often, multiple signalling pathways are active within the same cell to co-ordinate transcription of target genes (Figure 5.2). Due to this, the availability of downstream TFs within a cell is dependent on specific morphogen dose requirements being met. When they are not met, an embryo will fail to gastrulate correctly. *Wnt3*$^{-/-}$ murine embryos do not form a PS, mesoderm or node (Behringer *et al.*, 1999). Similarly, embryos that lack NODAL activity arrest before the onset of gastrulation and fail to form a primitive streak and A-P axis (Conlon *et al.*, 1994; Robertson *et al.*, 2003; Zhou *et al.*, 1993), whilst BMP4 null embryos fail to gastrulate due to a loss of primordial germ cells and mesoderm (Lawson *et al.*, 1999; Winnier *et al.*, 1995). Loss of the NODAL effector *Foxa2* results in embryos that lack a functional APS, resulting in somite fusion, and absent AME and severe anterior truncation (Ang and Rossant, 1994; Weinstein *et al.*, 1994).

The cells of the APS can be considered to form either the early gastrula organiser (EGO; 6.5 dpc) or the mid-gastrula organizer (MGO; 6.75 dpc), before the morphologically distinct node emerges. Briefly, the EGO consists of a group of 40 cells capable of inducing a secondary axis like that developed by the node, though murine fate mapping studies showed that the EGO on its own is not sufficient for anterior axial mesoderm generation (Robb and Tam, 2004; Tam and Steiner, 1999; Tam *et al.*, 1997). The EGO expresses multiple organiser related genes such as *Chrd* to inhibit BMP signalling in the anterior notochord, *Gsc* to promote PrCP development, and *Nodal* and its downstream effector *Foxa2* to promote a high NODAL environment (Ang *et al.*, 1993; Filosa *et al.*, 1997; Klingensmith *et al.*, 1999; Zhou *et al.*, 1993). EGO cells are incorporated into the PS and reallocated anteriorly, resulting in the formation of the MGO. The MGO contains precursors for the whole anterior axial mesoderm and much, but not all, of the notochord (Kinder *et al.*, 2001), and expresses many similar transcription factors and signalling components to the EGO. Unlike the EGO, however, *Nog* is also expressed to inhibit BMP signalling in the anterior notochord, and *T* is expressed to promote WNT signalling in the notochord (Bachiller *et al.*, 2000; Belo *et al.*, 1998, 1997; Kinder *et al.*, 2001; Pfister *et al.*, 2007). Additionally, signalling molecules such as BMP7, NODAL, WNT3a, WNT5a and WNT8 are expressed in the MGO (reviewed in Robb & Tam 2004).

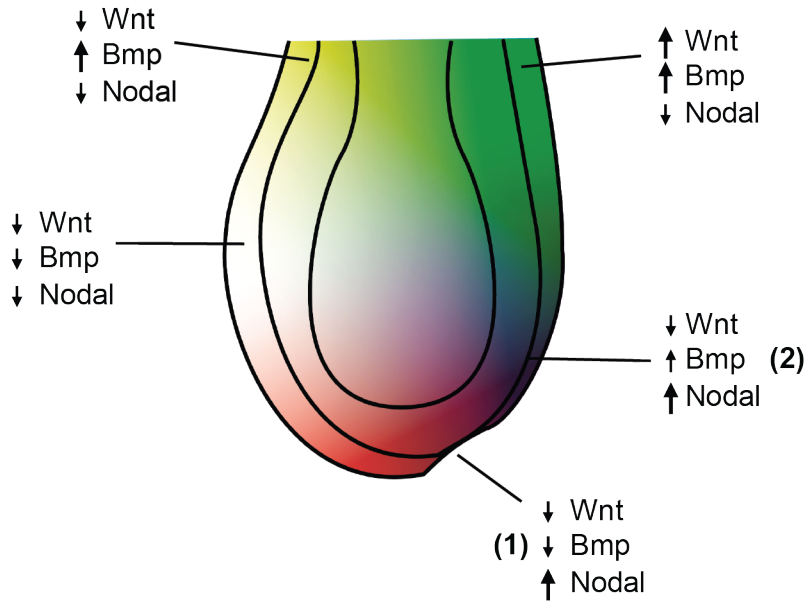


Figure 5.2: BMP, NODAL and WNT morphogen gradient activity overlaps in the developing embryos. When the morphogen gradients of NODAL (red), WNT (blue) and BMP (yellow) are overlayed in a late-streak stage embryo, they create pockets of high and low signalling environments. For example, **(1)** the node of a late-streak embryo exhibits medium-high NODAL signalling with low BMP and WNT signalling, whilst **(2)** the APS exhibits high NODAL signalling and medium-low levels of BMP and WNT signalling.

Finally, the node begins to form from the APS at the mid-streak stage (6.75 dpc) (Lee and Anderson, 2008; Yamanaka *et al.*, 2007) as described in Chapter 3. The node maintains medium-high levels of NODAL signalling and promotes expression of downstream NODAL targets (*Gsc* and *Foxa2*), along with low levels of both WNT and BMP activity (Figure 5.1-5.2) owing to the expression of their respective antagonists such as *Lhx1, Nog and Chrd* (Beddington 1994; Ang and Rossant 1994; reviewed in Davidson and Tam 2000; Brennan et al. 2002; Yamanaka et al. 2007). Despite this, WNT11 and BMP7 are active in the mid-gastrula node (Arkell and Beddington, 1997; Madabhushi and Lacy, 2011; Robb and Tam, 2004; Sinha *et al.*, 2015). The expression of WNT11 in this tissue thought to be due to a role in regulating the migration direction of cells from the APS through the node, as $Wnt5^{-/-}$; $Wnt11^{-/-}$ mutants exhibit reduced migration of PS stem cell descendants (Andre *et al.*, 2015; Cambray and Wilson, 2007). As the late-streak node is the presumed site of *ZIC2* requirement and its expression in this tissue is critical for the prevention of HPE, the medium-high NODAL environment, along with low WNT and BMP, will affect the types of TFs that are available to bind or interact with the *ZIC2* NCE. As such, any investigations into ability of the *ZIC2* NCE to drive gene expression must take into consideration the environment that its activity is being assayed in.

5.1.2 Model systems for studying developmental genes during gastrulation

The best approach to evaluate the pathogenicity of the *ZIC2* NCE variants is via direct manipulation of the mouse genome. This would assess causality, and enable assay of the direction of change (if any) in *Zic2* gene expression, but does not establish the mechanism by which the variants disrupt expression. Mechanistic studies would be assisted by the production of a cell-based model of NCE activity. Broadly speaking, two classes of cell lines could be considered for the production of such a model: primary cells lines derived from a relevant tissue (in this case, embryonic stem (ES) cells, epi-stem cells (EpiSC), or node derived primary cells), or transformed cell lines.

Cells that are taken from the inner cell mass of the mouse blastocyst (3.5 dpc) and cultured under appropriate conditions form undifferentiated ES cells (Evans and Kaufman, 1981; Martin, 1981). As these cells retain the capacity to repopulate the embryo and also contribute to all tissue of the adult, including the germline, they are considered naïve, pluripotent cells. (Bradley *et al.*, 1984; Nichols and Smith, 2009). In contrast, EpiSC are pluripotent cells that will give rise to all the embryonic germ layers (Brons *et al.*, 2007; Tesar *et al.*, 2007), but unlike ES cells they exhibit X-inactivation and cannot contribute to blastocyst chimeras (Guo *et al.*, 2009; Tesar *et al.*, 2007). ES and EpiSCs share multiple benefits, in that they are self-renewing and can thus be continuously propagated, can differentiate into most embryonic cell types under the correct conditions, and exhibit relatively good transfection rates (~85%) (Nichols and Smith, 2009;

Tamm *et al.*, 2016; Tesar *et al.*, 2007). Issues arise, however, when attempting to use either cell type to model specific embryonic structures such as the node. Differentiation of ES or EpiSC requires the coordination of multiple factors simultaneously to induce a target cell type. Previously, Winzi et al reported the generation of node/notochord like cells derived from murine ES cells, based on the expression of *Noto*. The creation of these cells first required induction of *Foxa2* and Brachyury-expressing progenitor cells via a low Activin A concentration, along with simultaneous inhibition of BMP, WNT and retinoic acid signalling in the presence of FGF2. The resulting cells, however, exhibited markers and morphology of both node and notochord populations. Additionally, this protocol required five days of differentiation before the node/notochord-like cells could be detected, with only 10% of the cell population reflecting the node/notochord state at day five (Winzi *et al.*, 2011). As such, it would be difficult to discern the mechanism of *ZIC2* NCE activity in the node via this method.

An alternative to ES and EpiSC is to use node-derived primary cells to analyse the activity of the *ZIC2* NCE. Primary cells espouse multiple benefits including their derivation directly from a tissue or organ, therefore they are the closest cell type to an *in vivo* model available. Additionally, primary cell lines have a relatively low mutation rate, therefore false interactions with the genes being investigated are unlikely to occur. Unfortunately, primary cell lines exhibit a finite life span, grow relatively slow, suffer changes with each passage and require supplemented media that is customised for their tissue type (Pan *et al.*, 2009). There is currently no available primary cell line derived from the mouse node.

Transformed cell lines, such as the human embryonic kidney (HEK293T) line, are a viable alternative and have several advantages. They are easily transfectable, which enables the rapid analysis of multiple conditions. It is often argued that transformed cell lines do not contain relevant tissue-specific proteins/processes and therefore provide little value for the study of tissue-specific mechanism. On the other hand, these cells can be considered 'test-tubes with a membrane' and ideal for assembling various mixtures of tissue-specific components via exogenous expression and/or culture in supplemented media.

The experiments in this chapter were conducted using the transformed cell line HEK293T which was originally derived from the kidneys of aborted human embryos with sheared adenovirus Ad(5) DNA, but is thought to potentially be derived from neural or adrenal cells within this population (Lin *et al.*, 2014; Shaw *et al.*, 2002). The cells were subsequently modified to express the SV40 Large T-antigen (DuBridge *et al.*, 1987), which interacts with the SV40 origin of replication present on the expression constructs used in this project to drive episomal replication to thousands of copy numbers per cell. As such, the effect of differences in transfection conditions (for example amount of DNA or efficiency) on overall expression levels

are minimised whilst protein expression and luciferase reporter activity are enhanced (Mahon, 2011). HEK293T cells express endogenous *ZIC2* and thus contain the fundamental transcriptional machinery for this gene (Hruz *et al.*, 2011). The cells however, do not express ZIC2 protein at high level (as judged by the inability of ZIC2 Antibodies to detect the protein in non-transformed cells). The Arkell laboratory routinely uses this cell line to conduct mechanistic studies (Ahmed *et al.*, 2013; Brown *et al.*, 2005; Pourebrahim *et al.*, 2011) and therefore protocols for the robust and reliable transient transfection of *ZIC* constructs already exist, as does knowledge regarding ways to switch on and off particular signalling pathways in this cell line. The HEK293T cell line was therefore selected for use in the below *ZIC2* NCE investigation.

5.1.3 Chapter 5 aims

As the *ZIC2* 3'UTR is predicted to contain an active ER element (termed the NCE) that drives expression of a target gene - most likely *ZIC2* - the aims of this chapter were:

- To determine the default activity of the *ZIC2* NCE (enhancer or repressor)
- To determine if the NCE is signal responsive, and if the six HPE-associated SNVs alter this response
- To determine whether any candidate TFs interact with the *ZIC2* NCE, and establish what effect, if any, the six HPE-associated SNVs have on this interaction.

5.2 Results

5.2.1 The *ZIC2* NCE acts as a repressor in basal culture conditions

To determine if the *ZIC2* NCE can drive reporter expression and thus act as an active ER element, heterologous reporter assays in HEK293T cells were employed. As NCEs are known to interact with promoters to enhance or repress transcription of their targets, a range of promoters were first tested with the *ZIC2* 3'UTR to determine which produced the highest levels of luciferase reporter transcription and greatest relative effect size. The human *β-globin* minimal promoter, amplified from plasmid pKS:β-globin:*lacZ* (Yee and Rigby 1993; Ahmed et. al., 2017, under review) and a *Hsp68* promoter (Hamada laboratory; Adachi et al. 1999) were chosen based on their proven ability to interact with NCEs and drive reporter expression in transgenic mouse models (Shiratori *et al.*, 2006). An untested putative human *ZIC3* promoter (Muenke laboratory, NHGRI, NIH) was chosen based on the knowledge that *ZIC3* and *ZIC2* share high levels of sequence conformation and regions of co-expression, therefore the *ZIC2* NCE would most likely be able to recognise and activate the *ZIC3* promoter. The wildtype *ZIC2* NCE (pCRW-8-TOPO-NCE-WT) was cloned into a promoterless pGL4.20-DEST construct directly, or pGL4.20-DEST constructs that contained either the *β-globin* minimal promoter (pβ), *Hsp68* promoter (pHsp68) or putative *ZIC3* promoter (pZ3) promoter via a Gateway LR reaction (Appendix Figure A3.3). Upon transfection of HEK293T cells with the resulting pGL4.20-NCE constructs (Appendix Figure A3.4), a luciferase assay was performed to detect reporter expression driven by the *ZIC2* NCE.

The presence of the wildtype *ZIC2* NCE significantly reduced the levels of luciferase reporter transcription (Figure 5.3 A and B) compared to the empty vector, irrespective of the promoter used, indicating the NCE's primary role in basal culture conditions is as a transcriptional repressor. Constructs containing pZ3 produced the highest levels of luminescence and largest fold difference between the empty vector and the WT *ZIC2* NCE compared to all other constructs tested. As pZ3 is likely the closest of the tested promoters to the endogenous *ZIC2* promoter and its inclusion in the constructs elicited the largest response, pGl-pZ3-DEST constructs were used for the remaining luciferase experiments in this chapter.

5.2.2 HPE-associated SNVs alter NCE transcription control

To determine whether the presence of each of the six HPE-associated SNVs affect NCE activity, each SNV was introduced into the NCE constructs with pZ3 and the luciferase assay in HEK293T cells repeated. The fold difference of each NCE construct relative to the empty pGl-pZ3-DEST is shown in Figure 5.3 C. Each HPE-associated SNV produced a significant increase in reporter activity compared to the wildtype NCE. For example, the M5 variant NCE produced

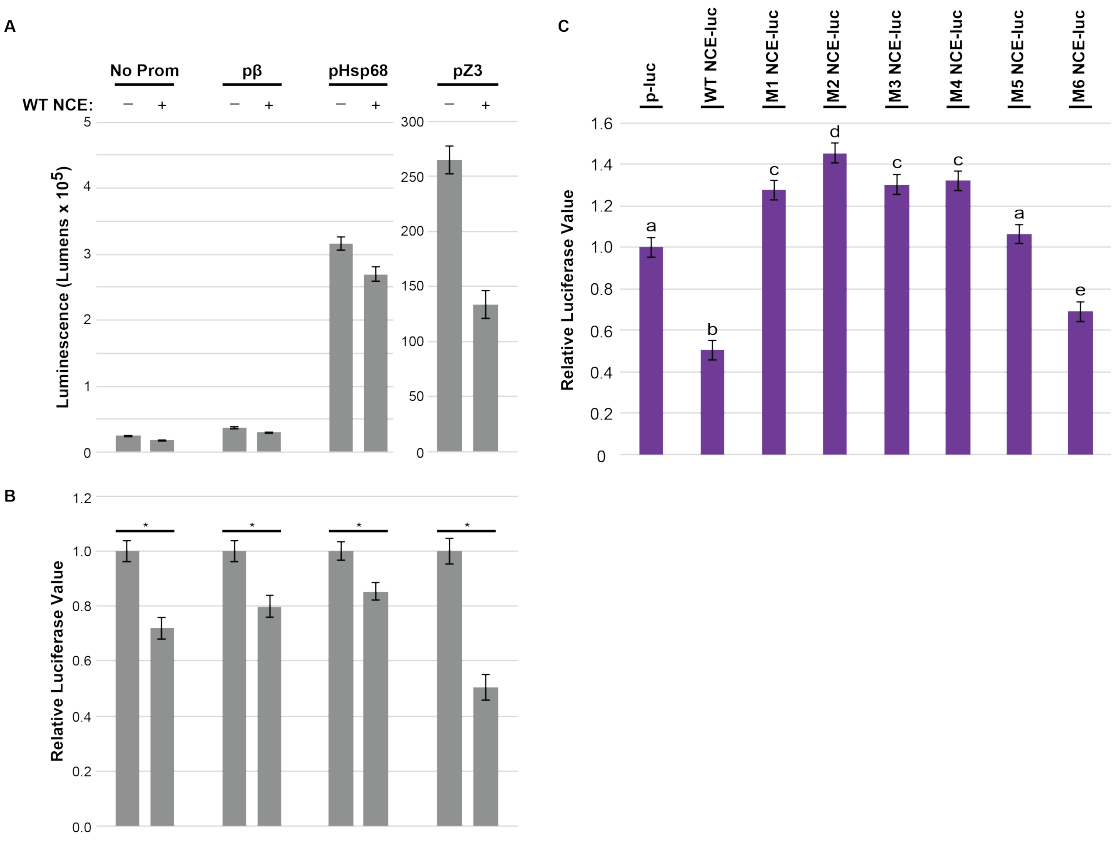


Figure 5.3: The ability of the WT *ZIC2* NCE to repress reporter activity is altered by the HPE-associated SNVs in basal culture conditions. (a-b) HEK293T cells were transfected with the reporter constructs pGL-DEST (No Prom), pGl-β-globin-DEST (pβ), pGl-Hsp68-DEST (pHsp68) and pGl-pZIC3-DEST (pZ3), with and without the WT *ZIC2* NCE. Comparisons of **(a)** luminescence levels and **(b)** fold values relative to the empty vector are shown. *: $p<0.05$, Student's T-Test. **(c)** HEK293T cells were transfected with the empty reporter construct pGl-pZic3-DEST (p-*luc*), pGl-pZIC3-WT *ZIC2* NCE (WT), and the mutated constructs pGl-pZic3-M1 *ZIC2* NCE to pGl-pZIC3-M6 *ZIC2* NCE (M1-M6). Reporter activity was measured as luminescence and fold values calculated relative to p-*luc*. Letters denote statistical significance of $p<0.05$ calculated via a two-way ANOVA with Fischer's unprotected post ad hoc test, where a is statistically significant to b and so on. Error bars = S.E.M of three external repeats.

relative values indistinguishable from background levels while the M2-4 NCEs all induced relative values significantly higher than background. In contrast, the M6 variant NCE retains its repressor function, although to a lesser degree than the wildtype NCE. These results indicate that each of the HPE-associated SNVs decreases the repressive effect of the NCE in basal culture conditions and, in the case of M1-4, converts the NCE to a weak enhancer in HEK293T cells.

5.2.3 The *ZIC2* NCE is signal responsive

To determine if the signalling environment of the cell determines NCE activity, HEK293T cells cultured in basal conditions were transfected with three reporter constructs: TOPFLASH (Helen Bellchambers, Arkell lab, unpublished), $[CAGA]_{12}$ and BRE, (Kumar et al. 2001; Morikawa et al. 2011; Korchynskyi & Ten Dijke 2002). Each reporter acts as a sensor for the level of the WNT, NODAL and BMP pathways, respectively. In unstimulated HEK293T cells, background luminescence levels of the BMP BRE reporter were considerably higher than the NODAL reporter $(CAGA)_{12}$ or the WNT reporter TOPFLASH, indicating that NODAL and WNT signalling are endogenously low in HEK293T cells, whilst BMP signalling is endogenously high (Figure 5.4 A and B, lanes with no ALK4 and NOGGIN added). Furthermore, transfection of HEK293T cells with the BMP receptors ALK2 and ALK6 failed to increase BMP signalling, again indicating that BMP signalling is endogenously high in HEK293T cells (Appendix Figure A3.6).

The level of activity of the pathways of interest can be modulated in various ways. For example, ALK4-HA transfection (to stimulate NODAL) (Figure 5.4), LiCl (to stimulate WNT [Pourebrahim et al. 2011]), V5-NOGGIN (to inhibit BMP, [Yuasa et al. 2005; Zimmerman et al. 1996]) (Appendix Figure A3.5), or a combination of ALK4-HA and V5-NOGGIN (to stimulate NODAL and inhibit BMP; Figure 5.4). Only the addition of ALK4-HA and a combination of ALK4-HA/V5-NOGGIN will be addressed in this chapter, as these signalling environments are most likely to be relevant to the region of the murine gastrula where the node develops. Transfection of the NODAL receptor ALK4 results in an increase of NODAL signalling as predicted, as well as a significant decrease in BMP and WNT signalling (Fig. 5.4, Table 5.2). WNT signalling was already low (based on small luminescence levels) and thus remained low with the overexpression of ALK4. BMP signalling, however, was deemed to be high in the absence of NODAL stimulation, thus the addition of ALK4 resulted in a reduction of BMP signalling by approximately half. The addition of ALK4 therefore results in the environment changing from a low NODAL, high BMP and low WNT environment to a high NODAL, medium BMP and low WNT environment (designated the 'red' environment). In contrast, the addition of the BMP inhibitor NOGGIN (which

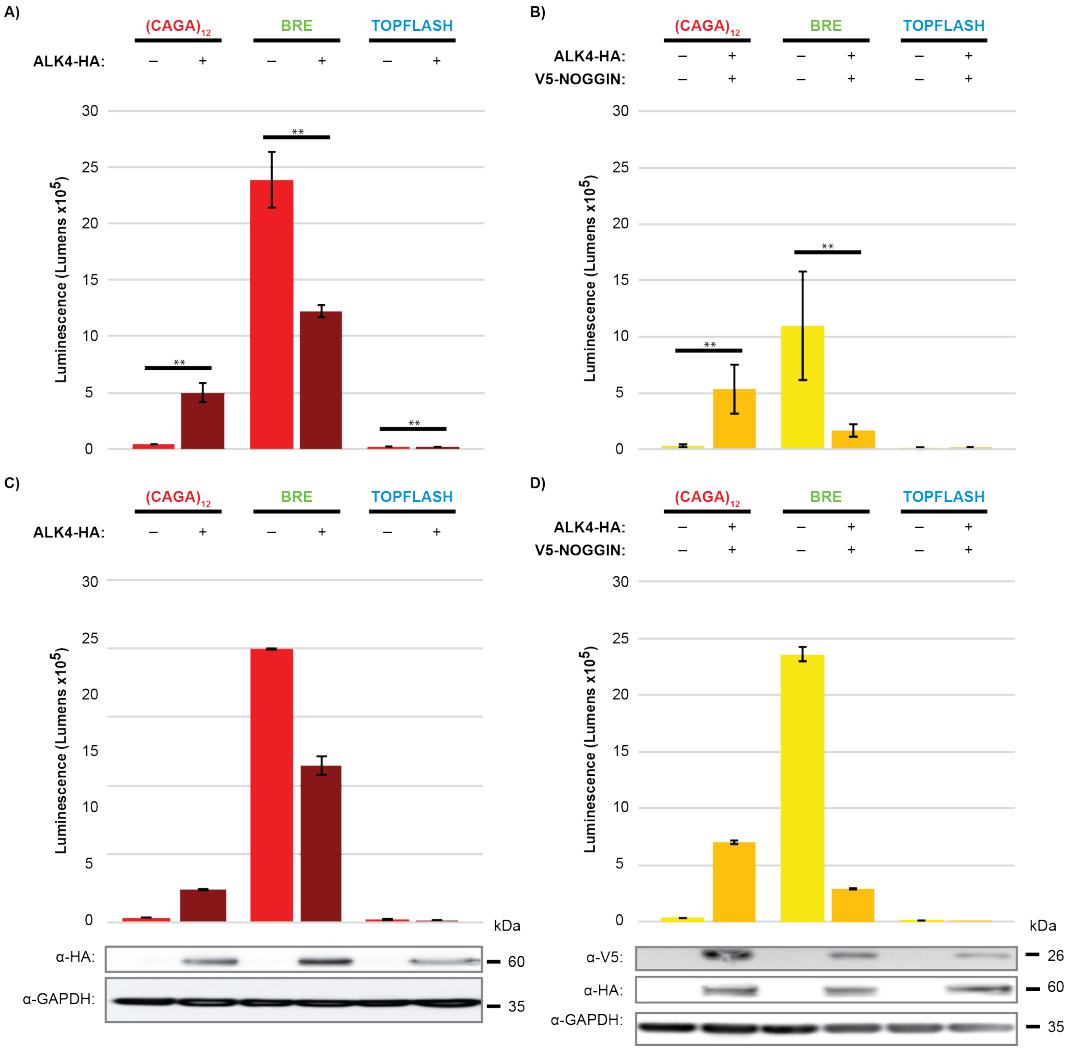


Figure 5.4: The signalling environment of HEK293T cells can be altered. HEK293T cells were transfected with the NODAL reporter $(CAGA)_{12}$, BMP reporter BRE or WNT reporter TOPFLASH, as well as ALK4-HA to stimulate NODAL signalling and V5-NOGGIN to inhibit BMP signalling, or an empty control vector (V5-DEST). Reporter activity was measured as luminescence. **(a)** The addition of ALK4-HA stimulates NODAL signalling in HEK293T cells, but decreases BMP and WNT signalling to create a high NODAL, medium BMP and low WNT environment. **(b)** The combined addition of ALK4-HA and V5-NOGGIN inhibits BMP and WNT signalling in HEK293T cells and stimulates NODAL signalling to create a high NODAL, low BMP and low WNT environment. Error bars = S.E.M of three external repeats; **: $p<0.01$, Student's T-Test. **(c-d)** Representative NODAL and BMP signalling experiments and western blots. In each case, a single representative experiment is shown, with error bars = S.D., N=3 internal repeats. Western blots correspond to the transfections shown. α-GAPDH blots are included as loading controls.

Table 5.1: Changes in HEK293T signalling activity via addition of ALK4-HA or a combination of ALK4-HA and V5-NOGGIN. Direction of increase/decrease is determined by directional changes in luminescence and relative luciferase activity direction shown in Figure 5.5, whilst quantitative amounts (small, medium and large) were determined by the comparative amounts of luminescence and relative luciferase activity for each pathway.

Signalling	**NODAL**	**BMP**	**WNT**
Unstimulated cells	Low	High	Low
'Red' environment	High	Medium	Low
	Large increase	Medium Decrease	Small Decrease
'Yellow' environment	High	Low	Low
	Large increase	Large decrease	Small Decrease

reduces BMP signalling in HEK293T cells [Appendix Figure A3.5]) in conjunction with ALK4 to induce high NODAL signalling results in the environment changing from a low NODAL, high BMP and low WNT environment to a high NODAL, low BMP and low WNT environment (designated the 'yellow' environment; Figure 5.4, Table 5.2).

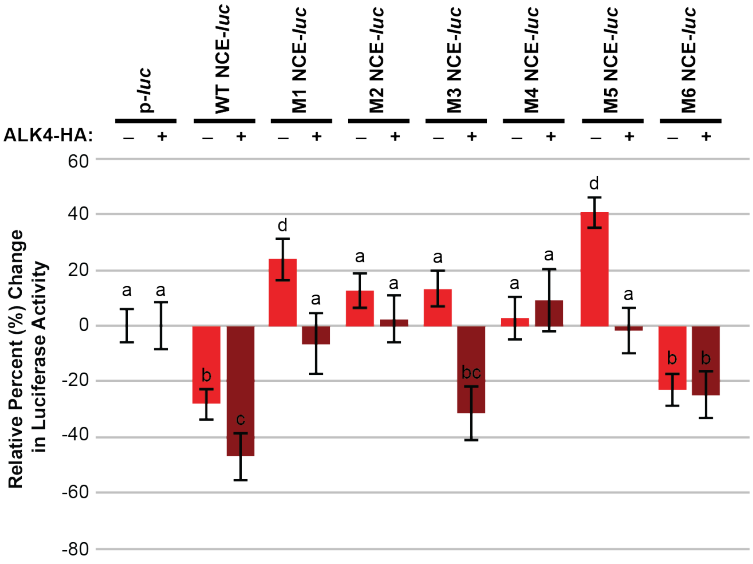
When the *ZIC2* NCE reporter constructs are transfected to either signalling environment (red or yellow), and the results normalised to account for the effect of the signalling environment on the construct's promoter activity (Appendix Figure A3.7), a change in luciferase activity is observed. In the 'red' signalling environment, the wildtype NCE shows further repression (Figure 5.8-5.9, Table 5.3). In contrast, repression is alleviated in the 'yellow' signalling environment ($p<0.05$). Together, this indicates that the NCE is a signal responsive repressor.

5.2.4 Mutation alters the NCEs response to the signalling environment

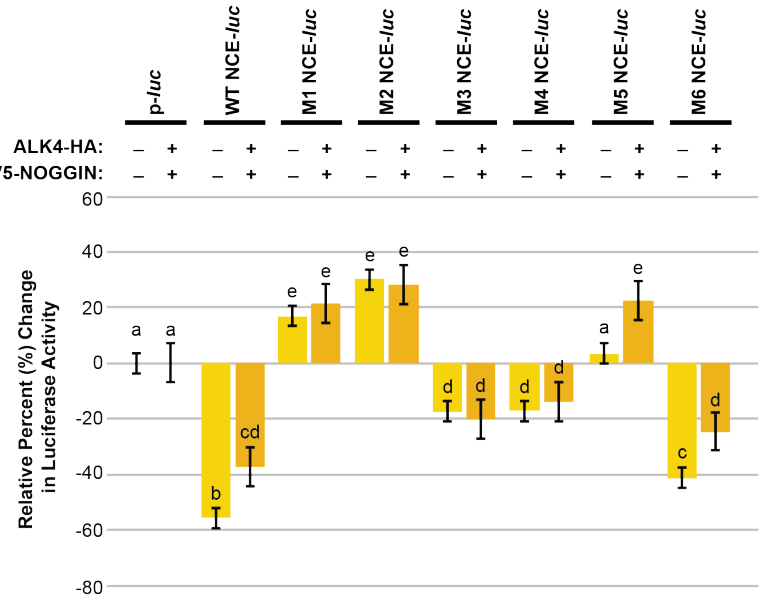
The effect of the six HPE-associated SNVs is altered by changes in signalling environment. For example, in a 'red' signalling environment (high NODAL, medium BMP, low WNT), the M1, M3 and M5 NCEs act as repressors, similar to the wildtype NCE (Figure 5.5 A, Figure 5.6 A). In contrast, the M2, M4 and M6 NCEs show no further repression, suggesting that the mutations disrupt the interactions of the NCE with a repressosome/enhanceosome responsive to high NODAL and medium BMP signalling. In contrast, a larger change in variant NCE activity when compared to the wildtype NCE can be seen in the 'yellow' signalling environment (high NODAL, low BMP, low WNT). Only the M5 and M6 NCEs alleviate repression in the same manner as the wildtype NCE under these signalling conditions, with the M1-M4 NCEs showing no such alleviation (Figure 5.5 B, Figure 5.6 B). This implies that a single SNV can alter NCE activity in a tissue-dependant manner in the embryo, depending on which signalling pathways are active in the tissue.

It should be noted that whilst the changes detected in this assay are statistically different, they also represent only a small effect size. This may be due to the use of a non-endogenous promoter, or the limited expression or absence of endogenous TFs in HEK293T cells that would normally interact with the NCE in embryonic tissues. Therefore, candidate TFs that are predicted to bind the *ZIC2* NCE should be added to the assay and analysed to determine whether a stronger effect can be elicited.

A



B



C

	WT NCE
'Red'	Further repression
'Yellow'	Alleviated repression

D

SNV	M1	M2	M3	M4	M5	M6
'Red'	Y	N	Y	N	Y	N
'Yellow'	N	N	N	N	Y	Y

Figure 5.5: The wildtype and mutated *ZIC2* NCE activity is signalling dependant. HEK293T cells were transfected with the empty reporter construct pGl-pZic3-DEST (p-*luc*), pGl-pZic3-WT *ZIC2* NCE (WT NCE-*luc*), and the mutated constructs pGl-pZic3-M1 *ZIC2* NCE to pGl-pZic3-M6 *ZIC2* NCE (M1-M6 NCE-*luc*), as well as **(a)** ALK4-HA to create a high NODAL, medium BMP and low WNT environment, **(b)** ALK4-HA and V5-NOGGIN to create a high NODAL, low BMP and low WNT environment, or an empty control vector (V5-DEST). Reporter activity was measured as luminescence. Fold values were calculated relative to p-*luc* and converted to the relative percent (%) change in activity compared to p-*luc,* whereby a value of 0-60 represents an increase in activity compared to background and -100-0 represents repression of activity. The activity of the **(c)** WT NCE and **(d)** six HPE-associated SNVs is summarised. 'Y' indicates that the mutated NCE responds to the addition of a TF in a similar manner to the wildtype NCE, whilst a 'N' indicates that the mutation responds differently. Error bars = S.E.M of three external repeats; letters denote statistical significance of $p<0.05$ calculated via a two-way ANOVA with Fischer's unprotected post ad hoc test, where a is statistically significant to b and so on.

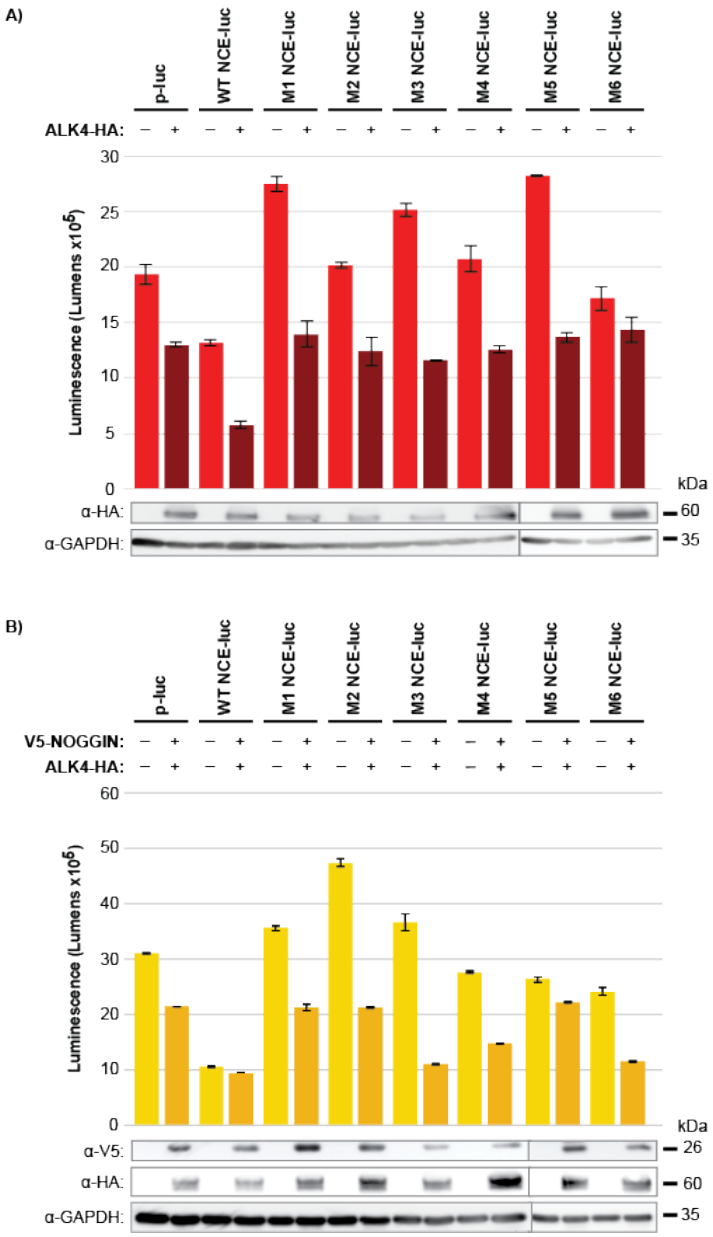
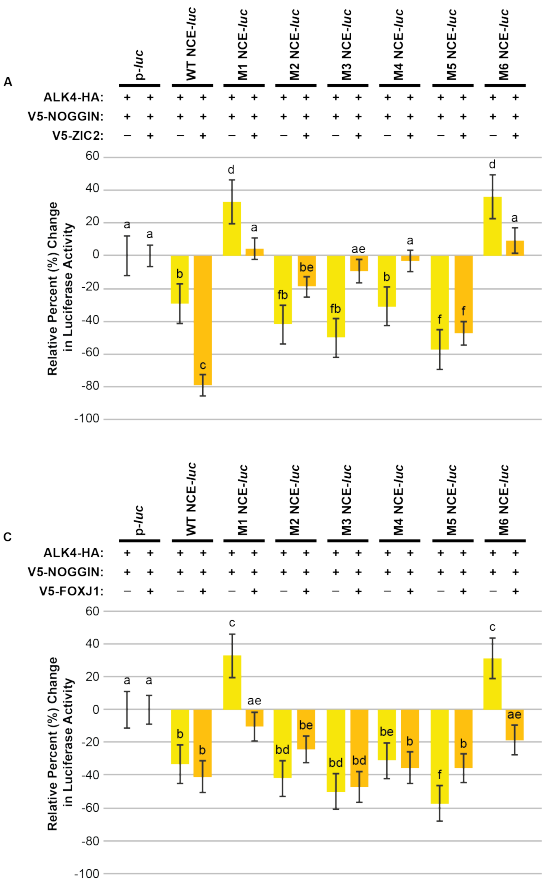


Figure 5.6: Representative experiment and western blot for the wildtype and mutant *ZIC2* NCEs in different signalling environments. HEK293T cells were transfected with the empty reporter construct pGl-pZic3-DEST (p-*luc*), pGl-pZic3-WT *ZIC2* NCE (WT NCE-*luc*), and the mutated constructs pGl-pZic3-M1 *ZIC2* NCE to pGl-pZic3-M6 *ZIC2* NCE (M1-M6 NCE-*luc*), in addition to either ALK4-HA to stimulate NODAL signalling, V5-NOGGIN to inhibit BMP signalling, or an empty control vector (V5-DEST). Reporter activity was measured as luminescence. In each case, a single representative experiment is shown, with error bars = S.D., N=3 internal repeats. Western blots correspond to the transfections shown. α-GAPDH blots are included as cytoplasmic loading controls. The western blot was conducted across multiple gels due to the large number of experimental samples.

5.2.5 Mutation alters the NCE's response to the overexpression of transcription factors

The activity of the WT and mutant *ZIC2* NCEs in the presence of overexpressed ZIC2, FOXA2 and FOXJ1 were assayed and the results normalised to remove any effect the added TFs had on the construct's promoter activity (Appendix Figure A3.8). In the 'yellow' signalling environment, overexpression of the TF *ZIC2* (a candidate to interact with the *ZIC2* NCE) augments the repressor activity of the WT NCE (Figure 5.7 A and D, Figure 5.8), suggesting that ZIC2 can interact with the NCE (either directly or indirectly) when in the presence of high NODAL signalling, but low BMP and WNT signalling. In the 'red' signalling environment, however, overexpression of ZIC2 elicits no response from the WT NCE (Figure 5.9 A, Figure 5.10), indicating that its interaction with the NCE is signal dependant. Overexpression of FOXA2 induces a similar response from the WT NCE in both signalling environments (Figure 5.7 B and D).

Similarly, the introduction of each HPE-associated SNV alters the NCE's response to overexpression of TFs. The ability of ZIC2 to induce further NCE repression in the 'yellow' signalling environments is retained with the introduction of the M1 and M6 variants, but lost with the introduction of M2-M5 variants (Figure 5.7 A and E). This suggests that M2-M5 interfere with ZIC2 directly binding to the NCE, or a protein complex containing (or downstream of) ZIC2 in these signalling conditions. The WT NCE is unresponsive to FOXJ1 overexpression in either signalling environment tested, however some NCE variants (except M2-M4 in either signalling environment) are capable of responding to FOXJ1 (Figure 5.7 C and E, Figure 5.8 C and E), suggesting the introduction of the mutations creates new FOXJ1 binding sites or binding sites for FOXJ1 co-factors. Moreover, the response from each variant NCE is dependent on the specific TF being overexpressed. M1, for example, responds to ZIC2 overexpression the same manner as the wildtype NCE, regardless of the signalling environment it is assayed in, but elicits a response different to the wildtype NCE when FOXA2 or FOXJ1 are overexpressed. Together, these results indicate that the effect of the six HPE-associated SNVs on NCE activity is not only signal dependant, but TF dependant as well.



D

TF	WT NCE
ZIC2	Further repression
FOXA2	Further repression
FOXJ1	No change

E

TF \ SNV	M1	M2	M3	M4	M5	M6
ZIC2	Y	N	N	N	N	Y
FOXA2	N	N	Y	N	N	N
FOXJ1	N	Y	Y	Y	N	N

Figure 5.7: The addition of candidate TFs alters the response of the wildtype and mutant *ZIC2* NCE in a high NODAL, low BMP and WNT signalling environment. HEK293T cells were transfected with the empty reporter construct pGl-pZic3-DEST (p-*luc*), pGl-pZic3-WT *ZIC2* NCE (WT NCE-*luc*), and the mutated constructs pGl-pZic3-M1 *ZIC2* NCE to pGl-pZic3-M6 *ZIC2* NCE (M1-M6 NCE-*luc*), as well as ALK4-HA to stimulate NODAL signalling and V5-NOGGIN to inhibit BMP signalling. **(a)** V5-ZIC2, **(b)** V5-FOXA2, **(c)** V5-FOXJ1 or an empty control vector (V5-DEST) were also transfected. Reporter activity was measured as luminescence. Fold values were calculated relative to p-*luc* and converted to the relative percent (%) change in activity compared to p-*luc,* whereby a value of 0-60 represent an increase in activity compared to background and -100-0 represents repression of activity. The activity of the **(d)** WT NCE and **(e)** six HPE-associated SNVs is summarised. 'Y' indicates that the mutated NCE responds to the addition of a TF in a similar manner to the wildtype NCE, whilst a 'N' indicates that the mutation responds differently. Error bars = S.E.M of three external repeats; letters denote statistical significance of $p<0.05$ calculated via a two-way ANOVA with Fischer's unprotected post ad hoc test, where a is statistically significant to b and so on.

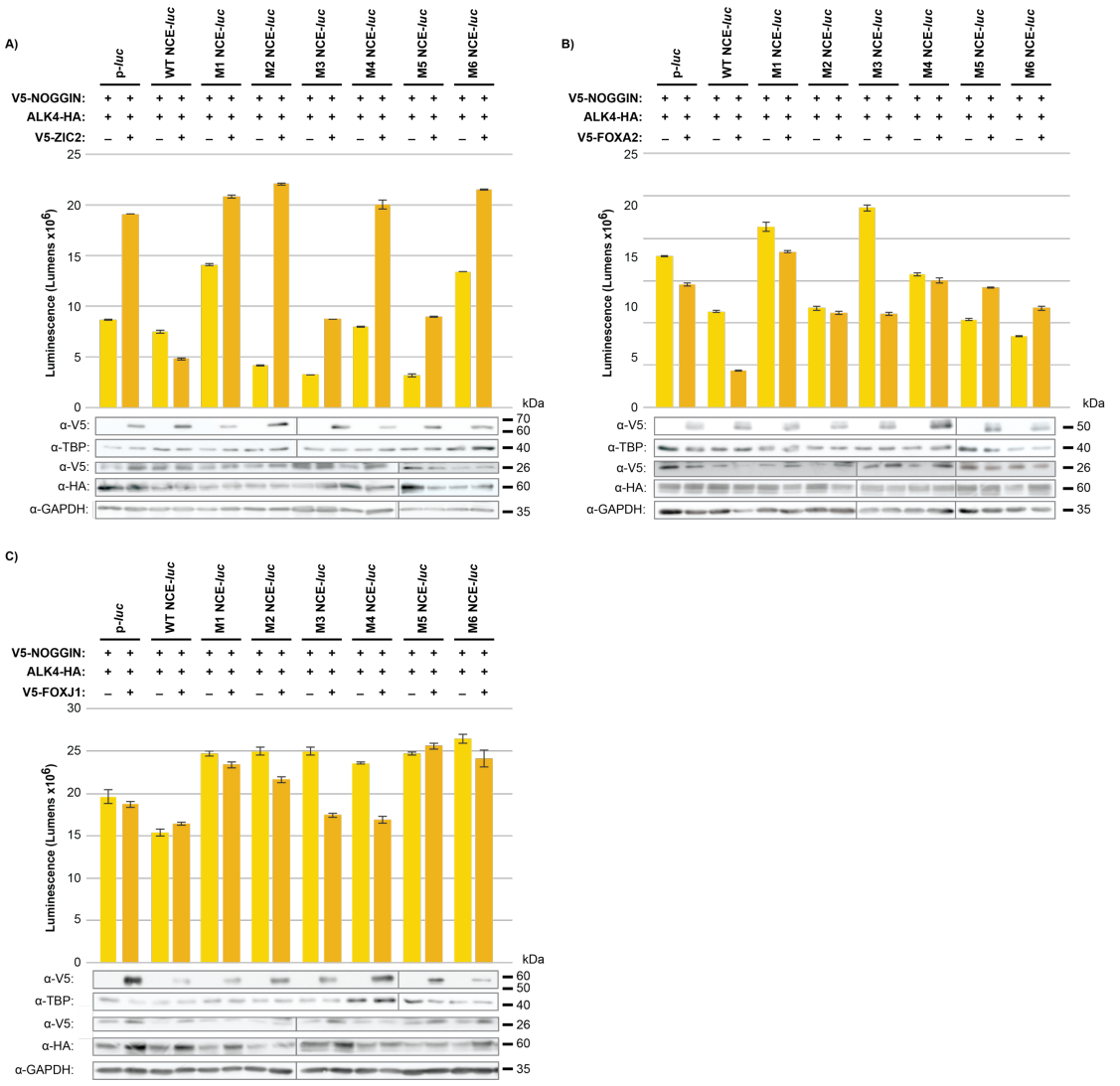


Figure 5.8: Representative experiment and western blots for the addition of TFs to the wildtype and mutant ZIC2 NCE in a high NODAL, low BMP and WNT signalling environment. HEK293T cells were transfected with the empty reporter construct pGl-pZic3-DEST (p-*luc*), pGl-pZic3-WT *ZIC2* NCE (WT NCE-*luc*), and the mutated constructs pGl-pZic3-M1 *ZIC2* NCE to pGl-pZic3-M6 *ZIC2* NCE (M1-M6 NCE-*luc*), in addition to ALK4-HA to stimulate NODAL signalling and V5-NOGGIN to inhibit BMP signalling, and **(A)** V5-ZIC2, **(B)** V5-FOXA2 or **(C)** V5-FOXJ1, or an empty control vector (V5-DEST). Reporter activity was measured as luminescence. In each case, a single representative experiment is shown, with error bars = S.D., N=3 internal repeats. Western blots correspond to the transfections shown. α-GAPDH (cytoplasmic) and α-TBP (nuclear) blots are included as loading controls. The western blots were conducted across multiple gels due to the large number of experimental samples.

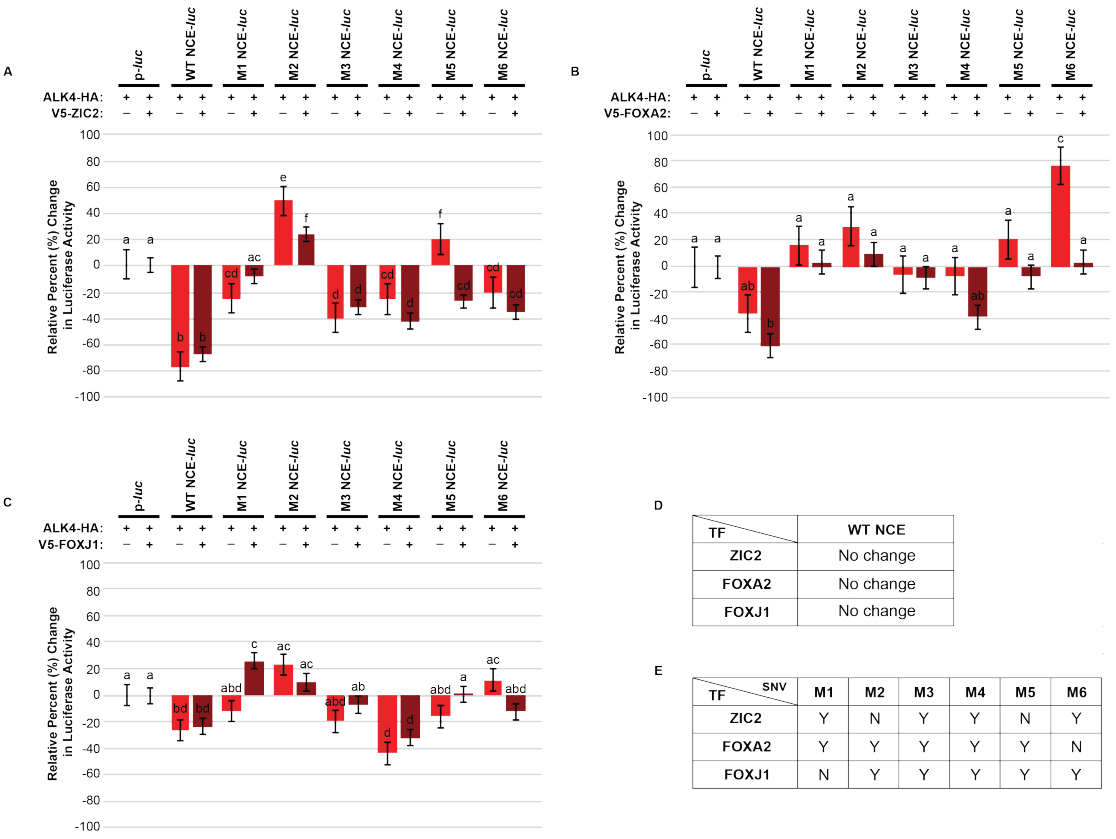


Figure 5.9: The addition of candidate TFs alters the response of the wildtype and mutant *ZIC2* NCE in a high NODAL, medium BMP and low WNT signalling environment. HEK293T cells were transfected with the empty reporter construct pGl-pZic3-DEST (p-*luc*), pGl-pZic3-WT *ZIC2* NCE (WT NCE-*luc*), and the mutated constructs pGl-pZic3-M1 *ZIC2* NCE to pGl-pZic3-M6 *ZIC2* NCE (M1-M6 NCE-*luc*), as well as ALK4-HA to stimulate NODAL signalling. **(a)** V5-ZIC2, **(b)** V5-FOXA2, **(c)** V5-FOXJ1 or an empty control vector (V5-DEST) were also transfected. Reporter activity was measured as luminescence. Fold values were calculated relative to p-*luc* and converted to the relative percent (%) change in activity compared to p-*luc,* whereby a value of 0-60 represent an increase in activity compared to background and -100-0 represents repression of activity. The activity of the **(d)** WT NCE and **(e)** six HPE-associated SNVs is summarised. 'Y' indicates that the mutated NCE responds to the addition of a TF in a similar manner to the wildtype NCE, whilst a 'N' indicates that the mutation responds differently. Error bars = S.E.M of three external repeats; letters denote statistical significance of $p<0.05$ calculated via a two-way ANOVA with Fischer's unprotected post ad hoc test, where a is statistically significant to b and so on.

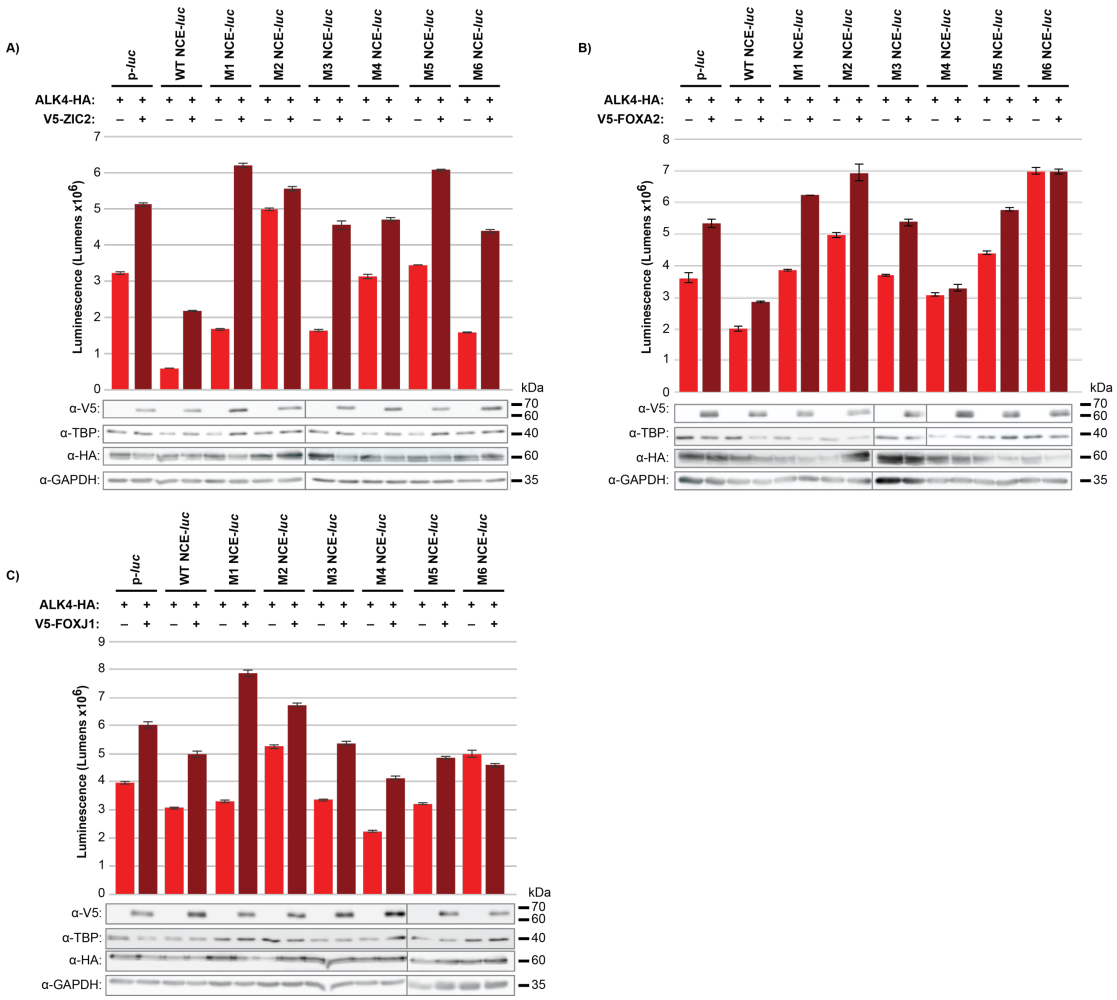


Figure 5.10: Representative experiment and western blots for the addition of TFs to the wildtype and mutant ZIC2 NCE in a high NODAL, medium BMP and low WNT signalling environment. HEK293T cells were transfected with the empty reporter construct pGl-pZic3-DEST (p-*luc*), pGl-pZic3-WT *ZIC2* NCE (WT NCE-*luc*), and the mutated constructs pGl-pZic3-M1 *ZIC2* NCE to pGl-pZic3-M6 *ZIC2* NCE (M1-M6 NCE-*luc*), in addition to ALK4-HA to stimulate NODAL signalling and **(A)** V5-ZIC2, **(B)** V5-FOXA2 or **(C)** V5-FOXJ1, or an empty control vector (V5-DEST). Reporter activity was measured as luminescence. In each case, a single representative experiment is shown, with error bars = S.D., N=3 internal repeats. Western blots correspond to the transfections shown. α-GAPDH (cytoplasmic) and α-TBP (nuclear) blots are included as loading controls. The western blots were conducted across multiple gels due to the large number of experimental samples.

When these interactions are correlated with the ZIC and FOX binding sites predicted to occur in the *ZIC2* NCE (Chapter 4), it can be seen that both the M4 and M5 SNVs occur within a binding site for the FOX family of transcription factors, whilst the introduction of M6 is predicted to create a new FOX binding site. It can therefore be hypothesised that any response seen in the M1-M3 NCEs from the overexpression of FOXA2 or FOXJ1 is most likely due to indirect (i.e. co-factor) interactions (as they are not predicted occur within FOX binding sites) whilst the responses elicited from M4-M6 NCEs is most likely due to direct interactions. As no mutations overlapped with a ZIC binding site, any responses elicited from overexpression of ZIC2 are most likely due to indirect interactions.

5.2.6 The signalling environments assayed correlate with in vivo signalling environments of the mouse embryo

To assess whether the manipulation of signalling pathways in HEK293T cells mimics gene expression in the node, the relative change in mRNA expression levels in endogenous target genes was measured. Embryos homozygous for the *Kumba* allele of *Zic2* exhibit decreased *Foxa2* and *Foxj1* expression in the node (Barratt *et al.*, 2014; Warr *et al.*, 2008), suggesting that ZIC2 functions upstream of both genes. On the other hand, overexpression of *zic2* in *Xenopus* embryos also results in a loss of *foxa2* expression in the dorsal blastopore lip (the equivalent tissue to the node) (Houtmeyers *et al.*, 2016), suggesting ZIC2 may also be required to limit expression of *Foxa2* in the node. It is possible that a negative feedback loop exists in which high levels of FOXA2 and/or ZIC2 represses *Zic2* transcription to prevent *Foxa2* overexpression.

HEK293T cells were transfected with V5-ZIC2, V5-FOXA2 or V5-DEST and exposed to both the 'red' (high NODAL, medium BMP and low WNT) and 'yellow' signalling environments (high NODAL, low BMP and WNT). In contrast to unstimulated HEK293T cells (Figure 5.7 A), the addition of ZIC2 into the 'red' signalling environment results in a significant decrease of *FOXA2* expression (Figure 5.7 B) but an increase in FOXJ1 expression, indicating this environment does not recapitulate known interactions between the three genes assayed. The addition of ZIC2 in a 'yellow' signalling environment results in a significant increase in endogenous *FOXA2* levels, and a small increase in *FOXJ1* (Figure 5.7 C). These trends correlate with ZIC2 positively regulating *FOXA2* and *FOXJ1* expression in the murine node and suggest that NCE activity in the 'yellow' signalling environment is an adequate model for this tissue. From these results, we can prediction the behaviour of the *ZIC2* NCE in this tissue. For example, the addition of ZIC2 results in a decrease in endogenous *ZIC2* expression and further repression of the ZIC2 NCE in the 'yellow' signalling environment, suggesting that ZIC2 uses the NCE to self-repress in the murine node.

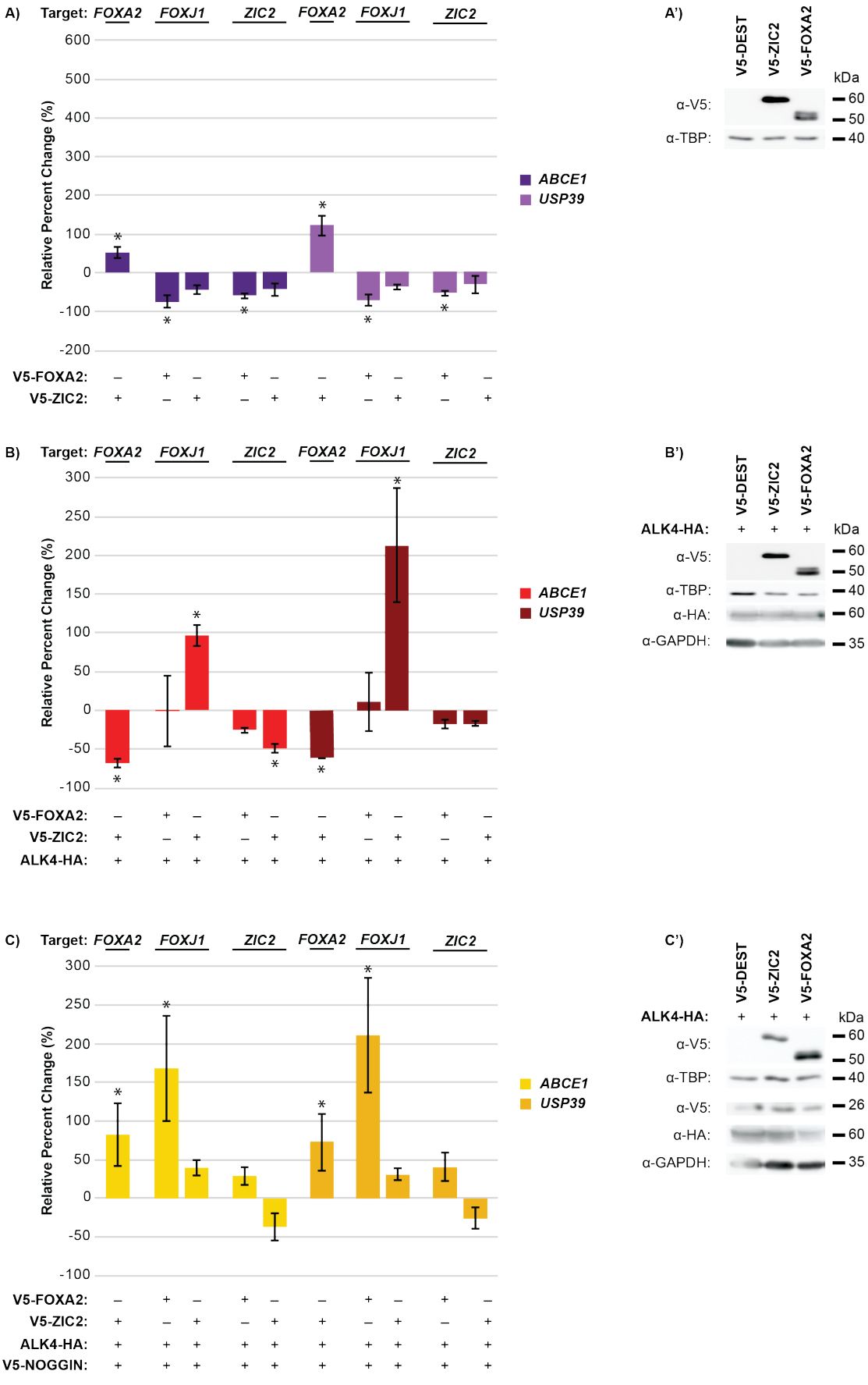

A)
Target: FOXA2 FOXJ1 ZIC2 FOXA2 FOXJ1 ZIC2
Relative Percent Change (%)
600
500
400
300
200
100
0
-100
-200
ABCE1
USP39
V5-FOXA2:
V5-ZIC2:
A')
V5-DEST
V5-ZIC2
V5-FOXA2
kDa
α-V5:
α-TBP:
60
50
40
B)
Target: FOXA2 FOXJ1 ZIC2 FOXA2 FOXJ1 ZIC2
Relative Percent Change (%)
300
250
200
150
100
50
0
-50
-100
ABCE1
USP39
V5-FOXA2:
V5-ZIC2:
ALK4-HA:
B')
V5-DEST
V5-ZIC2
V5-FOXA2
ALK4-HA:
kDa
α-V5:
α-TBP:
α-HA:
α-GAPDH:
60
50
40
60
35
C)
Target: FOXA2 FOXJ1 ZIC2 FOXA2 FOXJ1 ZIC2
Relative Percent Change (%)
300
250
200
150
100
50
0
-50
-100
ABCE1
USP39
V5-FOXA2:
V5-ZIC2:
ALK4-HA:
V5-NOGGIN:
C')
V5-DEST
V5-ZIC2
V5-FOXA2
ALK4-HA:
kDa
α-V5:
α-TBP:
α-V5:
α-HA:
α-GAPDH:
60
50
40
26
60
35

Figure 5.11: HEK293T cells can be induced to mimic the signalling environment of the embryonic node. HEK293T cells were transfected with V5-ZIC2, V5-FOXA2 or V5-FOXJ1 transcription factors. Results were normalised to *ABCE1* or *USP39* via the $\Delta\Delta^{CT}$ method, and percent change calculated relative to cells transfected with the empty vector V5-DEST to represent endogenous levels. A result of >0 indicates an increase in expression from endogenous levels, whilst <0 indicates a decrease in expression from endogenous levels. Error bars = S.E.M of three external repeats; * denote statistical significance of $p<0.05$ calculated via a one-way ANOVA with Fischer's unprotected post ad hoc test. **(A-C)** *FOXA2*, *FOXJ1* and *ZIC2* endogenous gene expression changes when transcription factors are transfected into **(A)** unstimulated (basal) HEK293T cells, **(B)** cells stimulated with ALK4-HA to create a high NODAL, medium BMP and low WNT signalling environment, or **(C)** cells stimulated with ALK4-HA and V5-NOGGIN to create a high NODAL, low BMP and low WNT signalling environment. **(A'-C')** A representative western blot, corresponding to the transfections and signalling environments shown. α-TBP (nuclear) and α-GAPDH (cytoplasmic) are included as loading controls.

5.3 Discussion

Multiple mechanisms combine to precisely control target gene transcription during embryo development. The binding of specific TF complexes to ER elements to either enhance or repress gene transcription is one such mechanism. The composition of these TF complexes is tightly regulated and subject to the signalling pathways that are active in the cell at a particular time. In this chapter I have shown that, regardless of signalling environment, the default role of the *ZIC2* NCE appears to be as repressive element and not an enhancer as first predicted in Chapter 4. The introduction of a single nucleotide mutation, however, is sufficient to significantly disrupt this repressive role. Additionally, I have shown that the candidate TFs ZIC2, FOXA2 and FOXJ1 interact with the wildtype and mutant *ZIC2* NCEs and alter reporter activity in specific signalling environments, providing direction for future investigations in the NCE and *Zic2*-associated HPE.

5.3.1 The wildtype *ZIC2* NCE is a repressor in HEK293T cells

The default role of the *ZIC2* NCE appears to be as a repressor element, as heterologous reporter assays under the control of the wildtype NCE consistently reported a reduction in luciferase activity when compared to background reporter activity, regardless of the signalling context. Moreover, the addition of TFs that are candidates to interact with the NCE (ZIC2, FOXA2 and FOXJ1) did not alleviate this repressive role. This result was unexpected, as the majority of identified ER elements function primarily as enhancers both in cell culture and in vivo. Though these elements can switch between enhancing and repressing roles in a context dependant manner, their default activity is as an enhancer. Indeed, as the role of enhancer elements in directing gene transcription was identified much earlier than that of repressors, the research community's knowledge on enhancers is far greater, and there are relative few examples of ER elements with a primary repressor role. The 3'UTR of *HMGA2*, a delayed early response gene expressed at the beginning of embryogenesis, is one such example. Upon transfecting HeLa cells with a construct containing the *HMGA2* 3'UTR situated 3' to a reporter gene, a 12.7-fold decrease in luciferase activity was detected. While initial deletion of the 3' end of the 3'UTR resulted in further repression, continued truncation of the 3' end resulted in a 20.5% increase in luciferase activity compared to the wildtype 3'UTR. This increase, however, remained lower than the background levels of reporter activity, suggesting the default role of the *HMGA2* 3'UTR was as a repressor element (Borrmann *et al.*, 2001). Borrmann concluded that the increase in reporter activity was due to an increase in transcript stability, as truncating the 3'UTR removed multiple AU-rich motifs contribute to mRNA decay. Removal of all the AU-rich motifs, however, still did not result in enhancer activity, suggesting that there remained a repressor element influencing reporter transcription as well as stability. Whilst the *HMGA2* 3'UTR luciferase results are similar to those achieved with the wildtype *ZIC2* NCE in this chapter, and the *ZIC2* 3'UTR is also predicted to contain AU-rich elements that influence transcript stability (Chapter 6), the

positioning of the *ZIC2* NCE upstream of the *Zic3* promoter and luciferase gene in the constructs used here ensure that luciferase output is solely due to the *ZIC2* NCE acting as an ER element and not due to it conferring transcript instability.

There remains the possibility, however, that the *ZIC2* NCE is made up of multiple positive and negative regulatory domains like those seen in the mouse *Albumin* ER element and the 3'UTR of *HGMA1* (Borrmann *et al.*, 2001; Herbst *et al.*, 1989). A series of truncated *ZIC2* NCE constructs can be created and put through the assays outlined in this chapter to determine if it is the entire NCE that confers repression onto a target reporter, or only specific domains. Analysis such as this would also allow for the prediction of how new HPE mutations in the *ZIC2* 3'UTR will affect the NCE, as those that occur in a repressive domain are more likely to result in an increase in target gene expression, whilst those that occur in a positive domain would presumably result in repression of a target. Additionally, one of the hallmarks of an active ER element is the ability to function independent of orientation. It remains to be determined whether the repressor function of the *ZIC2* NCE is maintained when the element is in the reverse orientation. Therefore, in addition to a series of truncated NCE constructs, constructs in which the NCE is reversed should also be investigated.

In vivo foot printing of the murine *Alb1* enhancer in gut endoderm found that sequential binding of TFs is required for function, rather than enhanceosome formation in one go (Gualdi *et al.*, 1996). Pioneering TFs such as FOXA2 were shown to initiate the sequential binding to enhancer elements, priming the element to be activated before other TFs were recruited to the element (Gualdi *et al.*, 1996; Zaret and Carroll, 2011). It should therefore be considered that addition of FOXA2 may prime the *ZIC2* NCE for the formation of an enhanceosome, but cannot recruit sufficient endogenous proteins to create a functioning enhanceosome, either due to under expression or complete lack of expression in HEK293T cells. This does seem unlikely, however, as this scenario would presumably lead to a small but detectable increase in activity upon addition of exogenous TFs. In contrast, addition of exogenous TFs consistently resulted in stronger repression of reporter activity by the *ZIC2* NCE. Nevertheless, transfection of multiple TFs into the above assays at the same time can be undertaken to determine whether a combination of proteins elicits a stronger response from the *ZIC2* NCE.

Due to time constraints, only three of the possible nine candidate TFs were tested for interactions with the NCE and. further investigation into the remaining candidate TFs identified in Chapter 4 is warranted. Both ZIC3 and ZIC5 are thought to act redundantly with ZIC2 in the embryonic tissues where they are co-expressed. Whilst this precludes the late-streak node as a possible site of interaction (as *Zic5* is not expressed in this tissue, and *Zic3* expression in the node occurs after *Zic2* expression has ceased [Chapter 4; Sutherland et al. 2013]), either ZIC could be

interacting with the *ZIC2* NCE in the PS and can therefore be tested in the APS-like signalling environment. CDX2 has been shown to maintain WNT signalling in the posterior of the mouse embryo, resulting in the generation of tissue required for trunk axial extension (Young *et al.*, 2009). When *Cdx2* is mutated, posterior truncations result (Chawengsaksophak *et al.*, 2004, 1997). In contrast, overexpression of the ZICs has been shown to inhibit WNT signalling in HEK293T cells and *Xenopus* embryos (Fujimi *et al.*, 2012; Pourebrahim *et al.*, 2011). A repressive interaction between ZIC2 and CDX2 is suggested by the inverse expression pattern of both genes in the PS, as *Zic2* expression recedes anteriorly down the streak at the same time that *Cdx2* begins to be expressed in the tissue that *Zic2* expression has ceased in.

5.3.2 The six individual HPE-associated SNVs alter NCE activity in a context-dependent manner

Multiple studies have demonstrated that SNVs in ER elements are able to disrupt the binding of TFs to the element. A G>A mutation in the IRF6 enhancer element disrupts the binding site for AP-2α, resulting in an increased incidence of cleft lip in probands (Rahimov *et al.*, 2008). Similarly, four distinct SNVs in the *Lmbr1* intronic enhancer disrupt *SHH* expression in the zone of polarising activity of the limb bud, resulting in preaxial polydactyly in both man and mouse (Lettice *et al.*, 2003), whilst SNVs in enhancers for *Tbx5* and *Sox9* have been shown to cause congenital heart defects and cleft palates, respectively (Benko *et al.*, 2009; Smemo *et al.*, 2012). The reduction in *ZIC2* NCE repressor activity to varying degrees upon the introduction of the six HPE-associated SNVs, independent of the signalling environment they were analysed in, agrees with the hypothesis that the SNVs are causative. If NCE function in stimulated HEK293T cells is faithful to NCE function *in vivo*, then alleviation of repression brought about by the SNVs is predicted to result in an increase in *Zic2* expression, which is the most likely target of the NCE. This increase of ZIC2 may occur in cells in which *Zic2* is normally required at a specific level, thus resulting in increased interaction with downstream targets, or exogenous expression in cells in which *Zic2* transcription is normally repressed. Though most *Zic2*-associated HPE cases are *Zic2* loss-of-function, it is probable that this gain of *Zic2* function will also lead to HPE. One published incidence of ZIC2 gain-of-function in human HPE is known. A Q36P mutation in the *ZIC2* CDS was found to result in a 170% increase in *ZIC2* transactivation activity in HEK293T cells when coupled to an ApoE promoter (Brown *et al.*, 2005). Though this mutation has not been definitively shown to cause a HPE phenotype on its own via mouse models, it provides precedent for ZIC2 gain-of-function as a mechanism for HPE development. Additional evidence comes from *zic2* knockdown and upregulation via morpholino and mRNA injections, respectively, in *Xenopus* embryos. These loss- and gain-of-function experiments both resulted in loss of *foxa2* expression in the

developing embryo and gastrulation defects (Houtmeyers *et al.*, 2016). It is therefore feasible that mutation of the *ZIC2* 3'UTR could lead to a gain of *Zic2* transcript, resulting in HPE.

The finding that each individual SNV confers different NCE activity in a context-dependant manner supports the hypothesis that the NCE is made up of multiple domains that work together to confer net regulation, rather than one total element. Moreover, the finding that each SNVs elicits a different response to overexpression of TFs predicted to bind the NCE suggests a mechanism is at play whereby the SNV alters the enhanceosome or repressosome that interacts with that specific region of the 3'UTR. Additionally, it suggests that some mutations would result in a more severe HPE phenotype than others. Indeed, those human probands in which the SNVs were identified exhibit the full spectrum of HPE defects (Chapter 1). As the M4-M6 SNVs are hypothesised to disrupt a direct interaction with FOXA2 or FOXJ1 (due to the presence of predicted FOX binding sites and altered NCE activity), this interaction should be confirmed with ChIP or EMSA analysis.

5.3.3 *ZIC2* NCE mouse models or CRISPR cell lines are required for further investigation

Two pertinent avenues of investigation that were not addressed in this chapter are (i) whether the *ZIC2* NCE directly controls expression of *ZIC2*, or a nearby gene such as *ZIC5*, at gastrulation and (ii) whether the individual HPE-associated SNVs are sufficient to cause a HPE phenotype. Indeed, I have shown that the mutations can disrupt the ability of the NCE to function in the same manner as the wildtype NCE, but this does not necessarily lead to a HPE phenotype. One way to test whether the *ZIC2* NCE specifically controls *ZIC2* expression is to investigate whether the NCE directly interacts with the *ZIC2* promoter. This can be achieved via chromosome conformation capture (3C) coupled with sequencing (4C), which would map genome-wide interactions with the *ZIC2* promoter (Sahlén *et al.*, 2015). Enrichment of the *ZIC2* NCE in this mapping would indicate that an interaction was occurring, and may also identify other ER elements that control *Zic2* expression. This method would still, however, only indicate that the promoter and NCE can interact, not that they do *in vivo*. The best way to investigate the role of the NCE in regulating Z*IC2* transcription during embryogenesis, and how the introduction of the six SNVs affects this regulation, is via the creation of transgenic embryos with a reporter under the control of the wildtype or mutant NCEs, and via direction mutation of the NCE via CRISPR-Cas9 mutagenesis.

5.3.4 Conclusion

ER elements have evolved as a method for gene expression modulation in response to signalling pathways and upstream effectors. This is particularly important for expression of *Zic2* in the node of the gastrulating embryo, where alterations in expression are known to lead to HPE. In a signalling environment that recapitulates aspects of node gene interactions (the 'yellow'

signalling environment), the repression of reporter transcription is enhanced by overexpression of the TFs ZIC2 and FOXA2. The introduction of each of the six SNVs disrupts the repressive role of the NCE, suggesting that an increase in the NCEs target, presumed to be *ZIC2*, may be sufficient to cause HPE.

Chapter 6: In silico and in vitro analysis of ZIC2 transcript stability

Optimization of the RT-qPCR protocol, cloning and decay analysis in this chapter were developed and performed in conjunction with Kathryn Dickson, an Honours student in the Arkell lab.

6.1 Introduction

Gene expression is a balance of transcription production and degradation. This is particularly relevant in transient biological processes such as embryo development where a gene requires rapid expression in a short period of time, followed by a large, controlled decrease in transcript levels. It is estimated that 40-50% of the changes in gene expression in response to cellular signals occurs due to regulation of mRNA stability (Garneau *et al.*, 2007). Transcript degradation occurs via multiple *cis*-acting regulatory regions (such as AU-rich elements [ARE], GU-rich Elements [GRES], ARE Binding Proteins [AREBPs], miRNAs and alternative polyadenylation). These elements work independently or cooperatively to confer transcript stability or promote destabilization, and determine transcript half-life in a context-dependant manner.

6.1.1 Transcript stability machinery

Eukaryotic mRNAs contain two main stability determinants – a 5' 7-methylguanosine cap (5' cap) and a 3' poly(A) tail (Figure 6.1). Both are incorporated into the mRNA strand during transcription (Garneau *et al.*, 2007) and both protect the transcript from degradation. Once the capped and polyadenylated transcripts are transported to the cytoplasm and P-bodies, the poly(A) tail and 5' cap act synergistically to initiate translation. The 5' cap binds the Translation Initiation Complex (TIC) whilst the poly(A) tail is bound by Poly(A) Binding Protein (PABP). Bound PABP also associates with the TIC, forming a closed loop and increasing the integrity of the transcript by protecting it from exonucleases (Garneau *et al.*, 2007; Kuersten and Goodwin, 2003; Weill *et al.*, 2012). The transcript is then either translated immediately via association with a ribosome, silenced and stored in P-bodies for translation at a later date, or decayed (Garneau *et al.*, 2007; Parker and Sheth, 2007). When decay is initiated, most eukaryotic transcripts go through a process of deadenylation to reduce the poly(A) tail ~25-2000 bp nascent mRNA down to ~80 bp (Figure 6.1). Deadenylation can be carried out by the activity of deadenylases PAN2-3, followed by the CCR4-NOT complex (Garneau *et al.*, 2007), or via the deadenylase Poly(A)-specific Ribonuclease (PARN). Both deadenylation pathways are mRNA specific and it is currently unknown why one is favoured over the other for individual transcripts (Garneau *et al.*, 2007; Wu and Brewer, 2012).

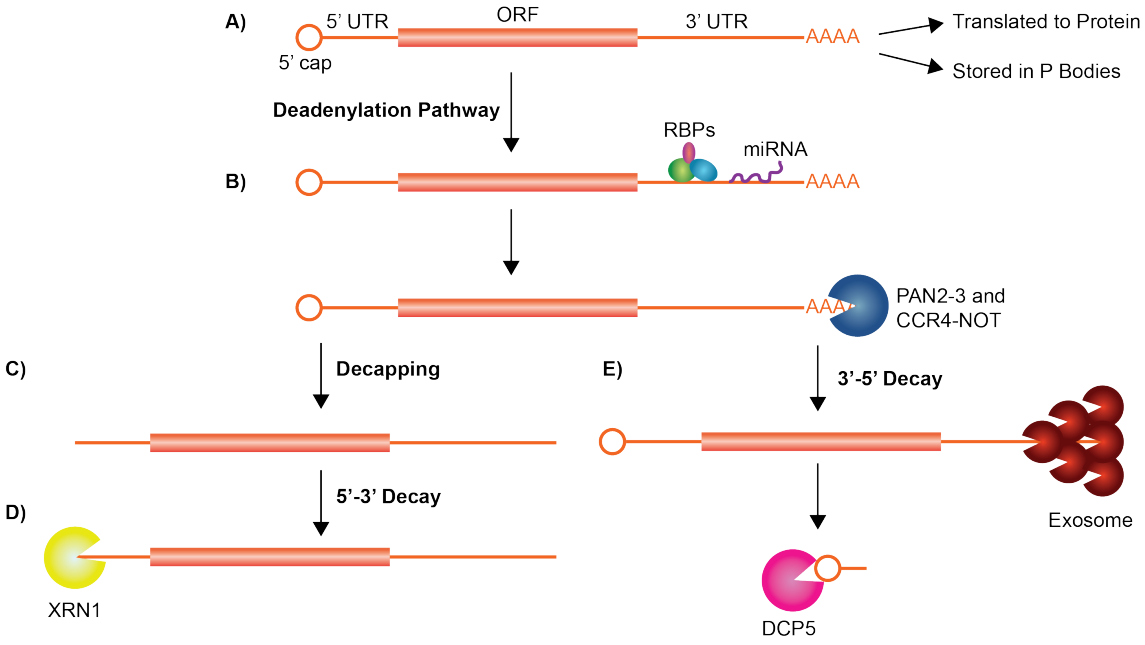


Figure 6.1: The eukaryotic mRNA decay pathway. **(a)** A 5' cap is added to the 5' end of all eukaryotic mRNA during the elongation step of transcription, where it confers protection from early degradation onto a transcript. Following cleavage of a transcript at a poly(A) site, a homopolymeric string of 25-2000 adenine nucleotides termed the poly(A) tail is added to the 3' end by Poly(A) Polymerase (PAP) (Garneau *et al.*, 2007). The poly(A) tail protects the mRNA from degradation, aids in the export of the mature mRNA to the cytoplasm, and is required for stabilizing transcripts to facilitate translation initiation (Garneau *et al.*, 2007; Weill *et al.*, 2012). The transcript is then either converted into protein, stored in P-bodies for later use or directed through a deadenylation pathway to be decayed. **(b)** A transcript is directed for decay when RBPs and miRNAs bind to recognition sites in the 3'UTR of a transcript, usually located in an ARE, and form an exosome. Deadenylases such as PAN2-3/CCR4-NOT or Poly(A)-specific Ribonuclease (PARN) bind to the poly(A) tail and trim the ~25-2000 bp nascent mRNA down to ~80 bp (Garneau *et al.*, 2007). This step is reversible. **(c)** The transcript is then either decapped (Brennan and Steitz, 2001) and **(d)** decayed in the 5-3 direction via exoribonucleases that bind to the newly exposed 5' end such as XRN1 (Garneau *et al.*, 2007; Wu and Brewer, 2012), or **(e)** the exposed 3' end is degraded via an exosome consisting of six proteins with homology to a 3'-5' phosphorolytic exoribonuclease and several accessory proteins. Once the body of the transcript is decayed, the remaining 5' cap is metabolized by the scavenger decapping enzyme DCP5 (Garneau *et al.*, 2007; Wu and Brewer, 2012). ORF: open reading frame, AAAA: poly(A) tail.

Following deadenylation, transcripts are either decayed in the 3'-5' direction via the scavenger decapping enzyme DCP5 (Figure 6.1), or the 5' cap is removed (Brennan and Steitz, 2001) and the transcript is decayed in the 5'-3' direction via the exoribonuclease XRN1. The choice of pathway is dependent on the individual transcript undergoing decay and the available machinery. Both the 5'-3' and 3'-5' pathways are involved in the decay of unstable ARE-containing mRNA transcripts in mammals (Garneau *et al.*, 2007; Wu and Brewer, 2012). A less prominent decay pathway termed Endoribunucleolytic Decay (ED) also occurs in eukaryotic cells. ED is initiated when an actively translating transcript is cleaved into two fragments that are susceptible to decay by XRN1 and an exosome (Garneau *et al.*, 2007; Wu and Brewer, 2012). Though this pathway is the most efficient method to destroy mRNA, it only occurs for specific genes and the transcripts must be actively translating (Garneau *et al.*, 2007).

6.1.2 In silico identification and analysis of transcript stability elements

Regions within 3'UTRs have been shown to impact the stability and decay of mRNA transcripts. These elements can be located in the 5'UTR, introns and coding regions of genes, but in the 3'UTR they are shielded from ribosomes and other translational machinery (Garneau *et al.*, 2007). To initiate the decay process, stability elements must first interact with the target transcript's 3'UTR.

AU- and GU-rich elements

AREs, defined by the pentamers 'AUUUA', range from 50-150 bp in length, are located within regions rich in adenine (A) and uridine (U) nucleotides (Chen and Shyu, 1995). Their predominant role is to target transcripts for degradation via binding by RNA binding proteins (RBPs) (Barreau *et al.*, 2005; Chen and Shyu, 1995; Danckwardt *et al.*, 2008). This ability was first demonstrated via fusion of the stable rabbit *β-globin* gene to an AU rich 3'UTR from granulocyte monocyte-colony stimulating factor (GM-CSF). The half-life of rabbit *β-globin* transcripts was reduced from 17 hours to less than 30 minutes in NIH3T3 derived cells upon the introduction of the 51 bp ARE region (Shaw and Kamen, 1986). While AREs typically promote transcript degradation, they can also provide a stabilising influence under the correct circumstances (Danckwardt *et al.*, 2008).

Investigations into ARE content have shown that flanking sequences effect the overall influence of the ARE on mRNA stability, and no two AREs are the same. A nonamer of 'UUAUUUAUU' or 'WWAUUUAWW' is the minimum sequence required for ARE function. Without the flanking bps, the sequence 'AUUUA' does not affect transcript stability (Chen and Shyu, 1995; Garneau *et al.*, 2007; Zubiaga *et al.*, 1995). Individual AREs can function independently, however multiple ARE regions in the one transcript produce an additive effect. Two AREs in the *cfos* 3'UTR were analysed via transient transfection of NIH3T3 cells with constructs containing the stable rabbit

β-globin gene and either each individual ARE or both combined. Increased transcript instability was conferred when both AREs were present, indicating that each element enhanced the destabilising ability of the other. Successive mutagenesis of each of three 'AUUUA' motifs in the *cfos* AREs led to a gradual loss of this destabilization (Chen *et al.*, 1994). The distance between pentamers can also affect transcript stability. The 3'UTR of the *PPP1R3* gene (encoding the RG1 protein involved in the balance of glycogen synthesis and glycogenesis) was found to contain a polymorphism in Pima Indians (Arizona) that removed 5 bp between two ARE pentamers. This led to lower transcript stability and decreased levels of RG1 protein, which coincides with a high level of insulin resistance and type 2 diabetes amongst Pima Indians compared to control populations (Hitti, 2012).

Approximately ~7% of human genes contain AREs, with the majority located in 3'UTRs. AREs are over represented in transcripts involved in cell proliferation, transcriptional regulation and developmental processes (Bakheet *et al.*, 2006; Garneau *et al.*, 2007; Halees *et al.*, 2008). There are three classes of AREs, with Class I and III elements generally found in transcripts of TFs and cell cycle regulatory protein and Class II AREs predominately found in the mRNA of cytokines (Wu and Brewer, 2012). Class I transcripts contain 1-3 scattered copies of the 'AUUUA' pentamer surrounded by predominately U rich regions, whilst Class II transcripts contain multiple overlapping copies of the pentamer in a U rich region. In contrast, Class III transcripts lack pentamers and instead contain only A and U rich sequences (Chen and Shyu, 1995; Halees *et al.*, 2008).

Recently, genome-wide microarray analysis of mRNA decay rates in primary human T cells identified a new subset of regulatory elements, termed GREs. Defined by the conserved 11 bp sequence of 'UGUUUGUUUGU', GREs are found in the 3'UTR of 11.98% of decaying transcripts that did not contain AREs or other known RNA regulatory motifs (Raghavan *et al.*, 2002; Vlasova *et al.*, 2008). In comparison, 9.90% of short-lived transcript contain a Class I ARE and 5.21% contain a Class II ARE (Vlasova *et al.*, 2008). Upon introduction of GREs from *c-jun*, *jun B* or *TNFRSF1B* into a stable *β-globin* reporter in primary human T cells, transcript destabilization occurred and decay rates increased from 53 hours to ~3 hours. Similar to AREs, the majority of transcripts containing GREs encode proteins involved in important regulatory processes such as TFs, proto-oncogenes and metabolism regulators (Vlasova *et al.*, 2008). Though relatively little is known about the behaviour of GREs, the RNA binding protein CUG-binding protein 1 (CEF1) is known to specifically bind GREs and mediate GRE-dependant mRNA decay (Vlasova-St. Louis and Bohjanen, 2011).

RNA Binding Proteins

At least 30 RBPs are known to bind to AREs (termed AREBPs) and promote or prevent destabilising RBPs from binding the transcript (Weill *et al.*, 2012). A shorter target transcript will have fewer AREBP binding sites available for regulation than a longer transcript. The majority of AREBPs act to recruit degradation machinery (such as deadenylases and exosome proteins) to the ARE to facilitate transcript decay (Garneau *et al.*, 2007). The best classified AREBPs are AU-rich element RNA binding protein 1 (AUF1), tristetraprolin (TTP), butyrate response factor 1 (BRF1), KH-type splicing regulatory protein (KSRP) and Hu proteins. Of these, AUF1, TTP, BRF1 and KSRP bind to AREs and promote the destabilisation of transcripts (Shaw and Kamen, 1986) by recruitment of the deadenylases CCR4-NOT and PARN (Weill *et al.*, 2012). AUF1 also prevents PABP from binding and protecting the transcript, thereby allowing the transcript's poly(A) tail to be deadenylated. Moreover, AUF1, KSRP and TTP recruit exosomes to AREs after the deadenylation process to promote further decay (Chen *et al.*, 2001; Loflin *et al.*, 1999). AREBPs can also confer stability by removing mRNA transcripts from decay sites or by competing with destabilizing factors for binding substrates, preventing decay. It is also hypothesised that stabilizing AREBPs are able to strengthen the interaction between PABP and the poly(A) tail, thereby preventing deadenylation from occurring (Garneau *et al.*, 2007). The Hu protein family, made up of HuR, HuB, HuC and HuD, is known to act as a stabilising factor when bound to AU rich sequences (Shaw & Kamen 1986; Fan & Steitz 1998b). HuR appears to complex with mRNA transcripts in the nucleus and protect them from degradation during transport to the cytoplasm, leading to an increase in gene expression (Peng et al. 1998; Fan & Steitz 1998).

microRNAs

miRNAs are ~21-26 bp long RNA molecules known to trigger endonucleolytic cleavage, promote repression of translation and accelerate decapping of transcripts, thereby silencing transcripts or promoting their decay (Valencia-Sanchez *et al.*, 2006). It is estimated that up to 30% of human genes are regulated by miRNAs (Valencia-Sanchez *et al.*, 2006), with half of the conserved motifs in 3'UTRs linked to miRNA binding (Lewis *et al.*, 2005). Suh *et al.* recently found that miRNAs are first required for mRNA regulation at 6.5 dpc and are supressed globally prior to this stage, with endogenous siRNA regulating these stages instead (Suh *et al.*, 2010). From 6.5 dpc onwards, miRNAs work to coordinate cell proliferation, differentiation and apoptosis, as well as play a role brain morphogenesis, stress resistance, fat metabolism and metabolic regulation (Brennecke *et al.*, 2003; Giraldez *et al.*, 2005; Wightman *et al.*, 1993; Xu *et al.*, 2003).

To promote transcript decay via miRNA, the AREBP TTP must interact with the Argonaut (Argo) family of RNA-induced silencing complex (RISC) component proteins, whilst the miRNA interacts with RISC itself. The TTP/RISC/miRNA complex that forms cooperates to recognise ARE

sequences in target transcripts which are bound by the miRNA (Jing *et al.*, 2005). Following binding, the complex sequesters target mRNA into P-bodies in the cytoplasm, increasing their association with decapping machinery but not exosome machinery, and results in 5'-3' direction decay or decay by an as yet unidentified pathway (Valencia-Sanchez *et al.*, 2006). Target mRNAs can also be stored in the cytoplasmic P-bodies where they will either be later degraded (Shyu et al. 2008; reviewed in Kulkarni et al. 2010) or returned to the pool of mRNAs yet to be translated (Shyu *et al.*, 2008). This storage is thought to be utilized during embryo development when a high level of mRNA is being made (Pillai *et al.*, 2005) and may play a part in the stringent temporal expression of particular signalling pathways.

3'UTRs that contain more than one ARE pentamer are able to interact with numerous different miRNAs at once, or with multiple copies of the same miRNA. This, coupled with differences in the strength of miRNA target binding, results in miRNA mediated degradation being transcript specific (Jing *et al.*, 2005). Additionally, as miRNAs are able to bind a transcript without having a perfectly complimentary sequence to the target, one type of miRNA is able to bind and regulate multiple transcripts at any one time (Baek *et al.*, 2008; Helwak *et al.*, 2013; Selbach *et al.*, 2008). If the miRNA is a complete complementary match to its target mRNA, the mRNA is cleaved and its protein levels are reduced or absent. If it is an imperfect match, the mRNA is destabilized and may still exhibit a reduction by an as yet unidentified mechanism (Fabian *et al.*, 2010). Because of this capacity to imperfectly bind a transcript, point mutations exert less influence on a miRNAs ability to recognise and decay a transcript than other decay elements such as RBPs. Whilst the introduction of polymorphisms or mutations may not prevent the miRNA from regulating its target mRNA, it will change the strength of the regulation. This has been demonstrated in *miR-189* and *SLITRK1* in humans with Tourette's syndrome, whereby a loss-of-function frameshift mutation results in shorter dendrites due to inhibition of miRNA-mediated mRNA repression (Abelson *et al.*, 2005).

Alternative polyadenylation

Cleavage of a transcript can occur at more than one PAS in a 3'UTR. More than half of mammalian transcripts contain multiple sites to allow for the creation of different length transcripts (Tian *et al.*, 2005). For cleavage to occur, a PAS must sit ~15-30 bp upstream of the cleavage site and a U/GU rich sequence must be ~30 bp downstream of the cleavage site (Beaudoing *et al.*, 2000; Weill *et al.*, 2012). Cleavage and polyadenylation specific factors such as CFI, CFII and PAP bind to the PAS in mammals, whilst other factors, such as Cleavage Stimulating Factor, bind to the downstream GU rich region (Beaudoing *et al.*, 2000). Together, the binding of these factors promotes cleavage of the transcript. Following cleavage, a poly(A) tail is added to the 3' end of the transcript (Weill *et al.*, 2012).

Whilst genes can contain multiple PAS in their 3'UTR, the most distal PAS is typically the default cleavage site. These PASs are commonly the canonical 'AAUAAA' or 'AUUAAA', which together make up 73.1% of all identified PAS's in the genome. In contrast, the proximal PAS's are predominantly made up of non-canonical sites, of which nine have so far been identified via analysis of human expressed sequence tags (ESTs) (Beaudoing *et al.*, 2000). The 3' default sites allows for longer transcripts to be created which undergo greater regulation, compared to those transcripts that use proximal PASs and contain fewer miRNA and RBP binding sites. This can be seen in genes required for embryonic development: those required in early development and in highly proliferative tissues are more likely to utilize multiple PAS to produce different sized transcripts, whilst those required for later development tend to have longer transcripts (Weill *et al.*, 2012). Alternative polyadenylation can also remove mRNA stability sites like AREs, which is practical for regulating gene expression by rapid transcript turnover. For example, the 3'UTR of *FGF2*, a gene required for embryo development and wound healing (Martín *et al.*, 2006; Ribatti *et al.*, 1999), makes up 90% of the mRNA transcript and contains eight PASs. When a destabilizing ARE that sits between the first and second poly(A) site is removed by alternative polyadenylation, transcript half-life changes from 110 minutes to 26-14 minutes, as determined by transfection of *in vitro* transcribed, capped and polyadenylated mRNAs and analysis of transcripts with a chloramphenicol acetyltransferase reporter in Cos-7 cells (Touriol *et al.*, 1999).

Mutations anywhere in the PAS region may disrupt poly(A) processing as successful polyadenylation requires a functional PAS, GU rich region and cleavage site (Beaudoing *et al.*, 2000). The majority of mutations found within a transcripts cleavage site or in the GU rich region are gain-of-function. In contrast, mutations that occur in or near the PAS are usually loss-of-function (Danckwardt *et al.*, 2008). For example, two thalassemias in humans have been attributed to mutations in the 'AAUAAA' PAS of both the α*-globin* and β*-globin* genes, resulting in inactive or severely inhibited gene expression and defective haemoglobin production (Harteveld *et al.*, 1994; Higgs *et al.*, 1983; Jankovic *et al.*, 1990; Orkin *et al.*, 1985).

6.1.3 Enhancer elements transcribe RNA

Enhancer RNAs (eRNAs) are short-lived nuclear RNAs, 50-2000 bp long on average, and transcribed from regions of enhancer elements devoid of H3K4me3 (Darrow and Chadwick, 2013; Kim *et al.*, 2010). eRNAs are commonly bidirectionally transcribed from an enhancer in low levels and contain a 5' cap (Djebali *et al.*, 2012; Kim *et al.*, 2010; Lam *et al.*, 2013), but in contrast to other RNAs, are not generally spliced or polyadenylated (Koch *et al.*, 2011). Only eRNA transcripts that are unidirectionally transcribed have been found to be polyadenylated, and these transcripts are usually >3 kb (Darrow and Chadwick, 2013; Kim *et al.*, 2010; Koch *et al.*, 2011). Though the transcription rates of eRNAs positively correlate to that of nearby protein

coding genes, eRNAs exhibit shorter half-lives than mRNAs and lncRNAs (Kim *et al.*, 2010; Lam *et al.*, 2013). Additionally, eRNA transcripts are enriched at active enhancers interacting with target promoters (Lin *et al.*, 2012; Sanyal *et al.*, 2012) and are suggested to be a better predictor of enhancer activity and identifier of target genes than conventional bioinformatics such as sequence conservation, chromatin modifications and TF binding (Andersson, 2014). The function of eRNAs is still under investigation, however targeted degradation of eRNAs via RNAi or RNase-H results in reduced expression of nearby protein coding genes in mouse macrophages (Lam *et al.*, 2013), suggesting eRNAs contain some functional role in gene regulation. Additionally, studies in multiple human and mouse cell lines have proposed a role for eRNA in stabilizing enhancer-promoter interactions and chromatin looping (Li et al. 2013), as well as recruitment of RNA Pol II to target promoters (Lam *et al.*, 2014).

6.1.4 Chapter 6 aims

- To annotate stability elements within the *ZIC2* 3'UTR.
- To assess the effect of the 3'UTR and NCE on transcript half-life.
- To assess the effect of the six HPE-associated SNVs on transcript half-life.

6.2 Results

6.2.1 The *ZIC2* 3'UTR contains AU-rich elements

Bioinformatic analyses can predict stability elements in 3'UTR sequences. When the human and mouse *Zic2* 3'UTR sequences were analysed via prediction algorithms at the AREsite database (Gruber *et al.*, 2011), a conserved 64 bp region (overlapping SNVs M3 and M4) that is predicted to be a Class I ARE was identified in both species (Figure 6.2). The motif has 'AUUUA' pentameters at both the 5' and 3', but only the 3' pentamer contains the known minimum flanking AU sequence required for ARE function (WWAUUUAWW). In support of this, the predicted *ZIC2* ARE sequence was significantly enriched in A and U nucleotides when compared to genomic 3'UTRs (p<0.001, two-tailed Chi-square, Table 6.1) and compared to the remaining nucleotides in the *ZIC2* 3'UTR. It is important to note, however, that the putative ARE may be longer than 64 bp as distal parameters of ARE regions are not well defined in the literature.

The presence of a GRE in the 3'UTR has been found to confer stability on a transcript similar to AREs. Thus, the *ZIC2* 3'UTR sequence was analysed for the presence of a GRE based on the conserved 'UGUUUGUUUGU' motif, allowing for one mismatch. The 11 bp motif was not found in either the human nor mouse *ZIC2* 3'UTR, suggesting that a GRE is not involved in *ZIC2* transcript degradation.

6.2.2 The *ZIC2* 3'UTR contains binding sites for RNA binding proteins and miRNAs

The RBPmap database (Paz *et al.*, 2014) and the RBPDB database (Berglund *et al.*, 2008) were used to identify predicted RBP binding sites in the *ZIC2* 3'UTR (Appendix Figure A4.1). Predictions were performed at 90% stringency and based on peer-reviewed and published data. Table 6.2 summarises the 40 RBPs predicted to bind the human *ZIC2* 3'UTR at 316 binding sites, and their binding frequency. Unlike the predicted TF binding sites in Chapter 4, the RBP binding sites are not enriched in the *ZIC2* conserved region (NCE) (Chi-squared goodness of fit test, p>0.05; Appendix Figure A4.1, Table 6.3). Furthermore, the predicted RBP sites do not cluster within the ARE. 3'UTR AREs are typically enriched in AREBP binding sites that regulate transcript decay or stabilization. Whilst there are 11 predicted binding sites for HuR within the *ZIC2* 3'UTR, none occur in the predicted ARE region. HuR is known promote transcript stability when bound to an ARE, and the absence of HuR binding sites suggests the *ZIC2* ARE most likely promotes transcript decay. It cannot be excluded that HuR can bind this region with an as yet unidentified binding site, however.

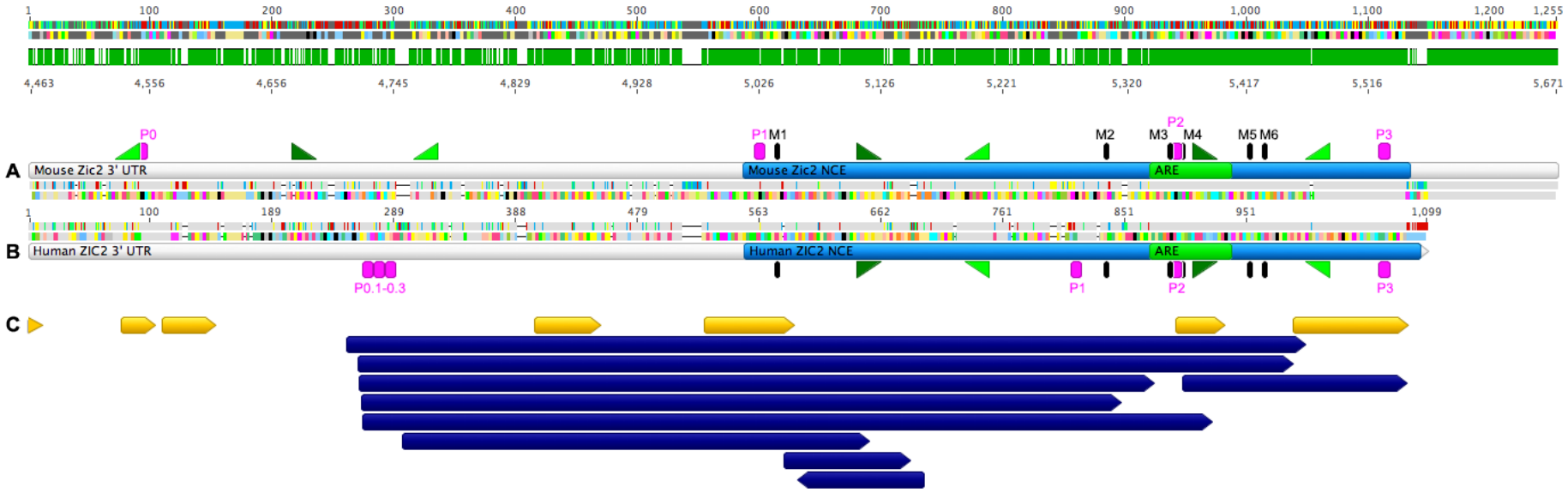

1
100
200
300
400
500
600
700
800
900
1,000
1,100
1,200
1,255
4,463
4,556
4,656
4,745
4,829
4,928
5,026
5,126
5,221
5,320
5,417
5,516
5,671
P0
P1
M1
M2
M3
P2
M4
M5
M6
P3
A
Mouse Zic2 3' UTR
Mouse Zic2 NCE
ARE
1
100
189
289
388
479
563
662
761
851
951
1,099
B
Human ZIC2 3' UTR
Human ZIC2 NCE
ARE
P0.1-0.3
P1
P2
P3
C

Figure 6.2: The *ZIC2* 3'UTR contains multiple predicted transcript stability elements. An alignment of the mouse and human *ZIC2* 3'UTRs. A 64 bp region was identified as a putative ARE based on the presence of two 'AUUUA' pentamers and a region high in A/U nucleotides in both the **(a)** mouse and **(b)** human *ZIC2* 3'UTR. This ARE region overlaps with the identified HPE-associated SNVs M3 and M4. Upon analysis for poly(A) sites, the murine *Zic2* 3'UTR was predicted to contain four poly(A) sites (designated mP0, mP1, mP2 and mP3, with P0 unlikely to be functional), whilst the human *ZIC2* 3'UTR was predicted to contain three poly(A) sites (designated hP1, hP2 and hP3). Sites P2 and P3 were conserved between the mouse and human 3'UTR, but P1 was not. **(c)** Long RNAs are predicted to be transcribed from within the *ZIC2* 3'UTR. CSHL Long RNA-seq (>200 nt; dark blue) and CSHL Small RNA-seq (<200 nt; yellow) (Cheng *et al.*, 2015; Langmead *et al.*, 2009; Parkhomchuk *et al.*, 2009) UCSC Genome Browser tracks were used to annotate RNA molecules predicted to be transcribed from within the *ZIC2* 3'UTR. It is possible that these RNA molecules represent eRNA. RNA-seq contigs with their 5' in the CDS of *Zic2* were excluded from analysis. Black, HPE-associated SNVs; pink, poly(A) sites; dark green triangles, forward oligonucleotides used in the poly(A) RT-qPCR assay; light green, reverse oligonucleotides used in the poly(A) RT-qPCR assay; light green rectangle, ARE.

Table 6.1: The *ZIC2* 3'UTR ARE is enriched in AU nucleotides. A two-tailed Chi-square was used to compare the AU and GC content in the *ZIC2* 3'UTR putative ARE (Observed) with the AU and GC content in 3'UTRs across the human genome (Expected) (L. Zhang *et al.*, 2004). A significance threshold of $p<0.05$ was used, with * indicating statistical significance**.**

	Nucleotides	**Observed**	**Expected**
***ZIC2* 3'UTR vs Genome**	AU	62.9%	57.6%
	GC	37.1%	42.4%
		Chi-square	1.026 with 1 degree of freedom
		p	=0.3110
***ZIC2* ARE vs *ZIC2* 3'UTR**	AU	79.7%	62.9%
	GC	20.3%	37.1%
		Chi-square	12.398 with 1 degree of freedom
		p	<0.001*
***ZIC2* ARE vs Genome**	AU	79.7%	57.6%
	GC	20.3%	42.4%
		Chi-square	19.869 with 1 degree of freedom
		p	<0.0001*

Table 6.2: The frequency of each RBP predicted to bind the *ZIC2* 3'UTR. Table correlates with Appendix Figure A4.1, which shows the positions of the binding sites. Frequency refers to the number of binding sites predicted to occur for each RBP. A * denotes RBPs with binding sites predicted to occur in the putative ARE.

RBP	Frequency	RBP	Frequency
A1CF	4*	PABPC	7
A2BP	1	PCBP	10
BRUNOL	1	PTBP	20*
CNOT	1*	PUM	12*
CPEB	8	QKI	1*
CUGBP	6	RALY	11
EIF4B	3*	RBM	14*
ELAVL	14	RBMX	8*
ESRP	1	SART	5
FUS	3	SFPQ	4
G3BP	2	SRSF	29*
HNRNP	21	TARDBP	12
HNRPLL	2	TRA	24
HUR	11	TUT	1*
IGF2BP	2	U2AF	7
KHDRBS	14*	YBX	3
KHSRP	2	YTHDC	1
MATR	2	ZC3H	8
MBNL	31*	ZCRB	3*
NOVA	5	ZNF	1

Enrichment of predicted RBP binding sites occurs at five out of the six HPE-associated SNVs nucleotides (p<0.01; Table 6.3), including M3 which resides within the ARE. In contrast, no putative binding sites overlap with M4, which also resides within the ARE. This suggests that select regions within the 64 bp ARE exhibit RBP binding site clusters, but not the entire element. Upon in silico introduction of the six identified SNVs, the binding sites of some RBPs, such as A1CF, were lost, whilst new binding sites were also created (Appendix Figure A4.1 and Table 6.4), including a site overlapping with M4. The SNVs therefore may change the content of the exosome bound to this region during the decay process and alter transcript stability.

Like RBPs, regulation of *ZIC2* transcript decay may occur via the binding of miRNAs to the *ZIC2* 3'UTR. The human *ZIC2* 3'UTR was screened for miRNA binding sites using the database miRanda (Betel *et al.*, 2008; John *et al.*, 2004), miRNASNP (Gong *et al.*, 2015) and DIANA-microT-CDS (Paraskevopoulou *et al.*, 2013; Reczko *et al.*, 2012). Predictions were performed at 90% stringency based on peer-reviewed and published ChIP data. Appendix Figure A4.2 and Table 6.5 detail the 248 miRNAs that are predicted to bind the *ZIC2* 3'UTR at 420 sites. Similar to RBPs, miRNA binding sites were not enriched in the NCE or ARE regions (p>0.05; Table 6.6), but were enriched at the six HPE-associated SNVs (p<0.01). Along with the lack of RBP binding site enrichment at the ARE, this suggests the entire 3'UTR is required to direct mRNA decay by binding of RBPs and miRNAs, rather than the ARE alone.

To determine what effect, if any, the six HPE-associated SNVs had on miRNA binding to the *ZIC2* 3'UTR, miRNASNP and mrSNP databases were utilised (Deveci *et al.*, 2014; Gong *et al.*, 2015). As miRNA are able to bind to their targets without a complete complimentary site, introducing the six SNVs into the 3'UTR did not create new or remove any existing miRNA binding site predictions. Their introduction, however, did result in a change in the predicted strength of the miRNAs binding to the *ZIC2* 3'UTR, as those sites with a higher level of complementation are more strongly bound by miRNA than those with lower complementation (Brennecke *et al.*, 2005). The introduction of M1 and M6 are predicted to result in an increase in binding strength for four miRNAs, whilst the introduction of M2 and M4 are predicted to result in a decrease in binding strength for three miRNAs (Table 6.7). These results suggest that the introduction of SNVs into the *ZIC2* 3'UTR may change the regulation of *ZIC2* transcripts by altering the ability of miRNAs to interact with the 3'UTR and direct decay.

Table 6.3: Predicted RBP binding sites are enriched at the HPE-associated SNVs in the wildtype *ZIC2* 3'UTR. A two-tailed Chi-squared goodness of fit test was performed to determine whether the predicted RBP binding sites were randomly distributed within the 3'UTR. Significance of $p<0.05$ (*) or $p<0.01$ (**) indicates predicted binding site enrichment in the region being tested. **(a)** Binding site enrichment in the non-coding conserved element (NCE; 540 bp) compared to the non-conserved region (559 bp) of the *ZIC2* 3'UTR. **(b)** Binding site enrichment in the AU-rich element (ARE; 64 bp) compared to the remaining *ZIC2* 3'UTR (1035 bp). **(c)** Binding site enrichment at the individual HPE-associated SNVs (1 bp) compared to the non-mutated 3'UTR (1089 bp). Observed and expected values are indicated. Degrees of freedom = 1 for all analyses.

A	Observed (n=316)	Expected	χ^2	*p*
Non-conserved region	153	160.74	0.759	0.3838
NCE	163	155.26		

B	Observed (n=316)	Expected	χ^2	*p*
Non-ARE	298	297.60	0.009	0.9234
ARE	18	18.40		

C	Observed (n=316)	Expected	χ^2	P
M1	3	0.29	25.348	0.0001**
M2	3	0.29	25.348	0.0001**
M3	3	0.29	25.348	0.0001**
M4	0	0.29	0.290	0.5900
M5	4	0.29	47.506	0.0001**
M6	4	0.29	47.506	0.0001**
Non-mutated 3'UTR	n – mutant observed	315.75		

Table 6.4: **RBP binding sites are lost or gained following introduction of the six HPE-associated identified SNVs.** Table correlates with Appendix Figure 4.1.

Mutation	Lost RBP	Gained RBP
M1	CPEB HNRNP MBNL	RBMX
M2	PCBP	HNRNP SRSF
M3	PUM A1CF	-
M4	-	A1CF
M5	KHDRBS PUM	PABPC QKI
M6	A1CF	-

Table 6.5: **The frequency of each miRNA predicted to bind the *ZIC2* 3'UTR.** Table correlates with Appendix Figure A4.2, which shows the positions of the binding sites. Frequency refers to the number of binding sites predicted to occur for each miRNA. A * denotes miRNA predicted to bind within the putative ARE.

miRNA	Freq.	miRNA	Freq.	miRNA	Freq.
hsa-let-7	2	hsa-miR-182	1	hsa-miR-520d-5p	3*
hsa-let-7a-2-3p	2	hsa-miR-185-5p	1	hsa-miR-522-3p	2*
hsa-let-7c-3p	1	hsa-miR-191-3p	1	hsa-miR-524-5p	3*
hsa-let-7g-3p	2	hsa-miR-204-3p	1	hsa-miR-539	1
hsa-miR-7-1-3p	1	hsa-miR-205	1	hsa-miR-539-3p	1
hsa-miR-7-2-3p	1	hsa-miR-212	1	hsa-miR-543	5*
hsa-miR-10b-3p	1	hsa-miR-224-3p	2*	hsa-miR-548	4*
hsa-miR-18a-5p	1	hsa-miR-297	1	hsa-miR-548aa	3
hsa-miR-18b	1	hsa-miR-302a-5p	2	hsa-miR-548aj-3p	5
hsa-miR-23a-3p	3	hsa-miR-320a	3	hsa-miR-548aj-5p	1
hsa-miR-23b-3p	3	hsa-miR-320b	3	hsa-miR-548ap-3p	3
hsa-miR-23c	3	hsa-miR-320c	3	hsa-miR-548ar-3p	4
hsa-miR-30a	1	hsa-miR-320d	3	hsa-miR-548as-3p	3
hsa-miR-30c	1	hsa-miR-323a	1	hsa-miR-548aw	1
hsa-miR-30d	1	hsa-miR-323a-3p	3	hsa-miR-548c-3p	5
hsa-miR-30e	1	hsa-miR-335-3p	4	hsa-miR-548e-5p	3
hsa-miR-32	2*	hsa-miR-337	1	hsa-miR-548f-5p	1
hsa-miR-33a	1	hsa-miR-340-5p	1*	hsa-miR-548g-5p	1
hsa-miR-33b	1	hsa-miR-371b-5p	1	hsa-miR-548t-3p	3
hsa-miR-92a-2-5p	1	hsa-miR-373-5p	1	hsa-miR-548u	1
hsa-miR-96-5p	1	hsa-miR-374b-3p	1	hsa-miR-548x-3p	5
hsa-miR-105-5p	3*	hsa-miR-374c	2*	hsa-miR-548x-5p	1
hsa-miR-125	1	hsa-miR-377-3p	3	hsa-miR-564	2
hsa-miR-125b-2-3p	3	hsa-miR-421	1	hsa-miR-582	1
hsa-miR-126-5p	1	hsa-miR-451b	1*	hsa-miR-583	2
hsa-miR-129-5p	2	hsa-miR-485-3p	1	hsa-miR-586	2
hsa-miR-1292-5p	1	hsa-miR-491	3	hsa-miR-590-3p	2
hsa-miR-1305	2	hsa-miR-491-5p	1	hsa-miR-607	2
hsa-miR-132	1	hsa-miR-493-5p	2*	hsa-miR-608	1
hsa-miR-1321	1	hsa-miR-494	2	hsa-miR-609	3
hsa-miR-133a-5p	1	hsa-miR-494-3p	3	hsa-miR-616-5p	1
hsa-miR-136-3p	1	hsa-miR-497	1	hsa-miR-620	1
hsa-miR-138	1	hsa-miR-497-3p	3	hsa-miR-624-3p	1
hsa-miR-140-3p	1	hsa-miR-499	2	hsa-miR-639	1
hsa-miR-145	1	hsa-miR-499a-3p	2	hsa-miR-655	3*
hsa-miR-149-3p	1	hsa-miR-499b-3p	2	hsa-miR-659-3p	1
hsa-miR-153-5p	4	hsa-miR-500a-5p	1	hsa-miR-663a	1
hsa-miR-181	1	hsa-miR-503-3p	1	hsa-miR-675-3p	2
hsa-miR-181a	2*	hsa-miR-513	2	hsa-miR-765	2
hsa-miR-181b	2*	hsa-miR-513b	3	hsa-miR-766-5p	1
hsa-miR-181c	2*	hsa-miR-513b-5p	3	hsa-miR-873-5p	1
hsa-miR-181d	2*	hsa-miR-5196-5p	1	hsa-miR-876	1

miRNA	Freq.	miRNA	Freq.	miRNA	Freq.
hsa-miR-876-3p	1	hsa-miR-3928	1	hsa-miR-4749-5p	2
hsa-miR-1237-3p	1	hsa-miR-4253	1	hsa-miR-4756-5p	1
hsa-miR-1238	2	hsa-miR-4262	3	hsa-miR-4761-5p	1
hsa-miR-1238-3p	1*	hsa-miR-4270	2	hsa-miR-4766-5p	2
hsa-miR-1246	2*	hsa-miR-4282	1	hsa-miR-4775	2*
hsa-miR-1252	1	hsa-miR-4306	1	hsa-miR-4784	1
hsa-miR-1255b-2-3p	1	hsa-miR-4311	2	hsa-miR-4788	1
hsa-miR-1261	1*	hsa-miR-4419a	1	hsa-miR-4789-5p	1
hsa-miR-1270	1	hsa-miR-4422	3	hsa-miR-5008-5p	1
hsa-miR-1271-5p	1	hsa-miR-4423-5p	1	hsa-miR-5582-3p	3
hsa-miR-1277-5p	2*	hsa-miR-4429	3	hsa-miR-5583-3p	1*
hsa-miR-1284	1	hsa-miR-4432	1	hsa-miR-5680	3
hsa-miR-1289	1	hsa-miR-4441	2	hsa-miR-5688	1
hsa-miR-1292-5p	1	hsa-miR-4446-3p	1	hsa-miR-6071	1
hsa-miR-1305	2	hsa-miR-4457	3	hsa-miR-6074	1
hsa-miR-1321	1	hsa-miR-4463	1	hsa-miR-6125	2
hsa-miR-1323	2	hsa-miR-4464	1	hsa-miR-6127	1
hsa-miR-1908	1	hsa-miR-4476	1	hsa-miR-6129	1
hsa-miR-3115	1	hsa-miR-4485	1	hsa-miR-6130	1
hsa-miR-3137	1*	hsa-miR-4490	1	hsa-miR-6133	1
hsa-miR-3148	1	hsa-miR-4495	2	hsa-miR-6504-3p	3*
hsa-miR-3150b-3p	1	hsa-miR-4633	2	hsa-miR-6510-5p	1
hsa-miR-3152-5p	1	hsa-miR-4633-5p	1	hsa-miR-6515-3p	1
hsa-miR-3153	1	hsa-miR-4644	1	hsa-miR-6715a-3p	2
hsa-miR-3154	5	hsa-miR-4646-5p	1	hsa-miR-6715b-3p	1
hsa-miR-3162-5p	1	hsa-miR-4651	1	hsa-miR-6748-5p	2
hsa-miR-3163	1	hsa-miR-4668-3p	2	hsa-miR-6754-5p	2
hsa-miR-3173-3p	1	hsa-miR-4670-3p	1	hsa-miR-6756-5p	3
hsa-miR-3177-3p	1	hsa-miR-4672	1	hsa-miR-6759-5p	2
hsa-miR-3179	1	hsa-miR-4694-3p	5*	hsa-miR-6766-5p	2
hsa-miR-3180	1	hsa-miR-4699-3p	2*	hsa-miR-6776-3p	2
hsa-miR-3196	1	hsa-miR-4706	2	hsa-miR-6780a-3p	1*
hsa-miR-3200-5p	1	hsa-miR-4719	3	hsa-miR-6793-5p	2
hsa-miR-3202	4	hsa-miR-4721	1	hsa-miR-6796-5p	1
hsa-miR-3622b-5p	1	hsa-miR-4728-5p	1	hsa-miR-6835-3p	2
hsa-miR-3650	1	hsa-miR-4731-5p	1	hsa-miR-6839-3p	1
hsa-miR-3658	1	hsa-miR-4734	1	hsa-miR-6875-3p	2*
hsa-miR-3671	1	hsa-miR-4735-3p	1	hsa-miR-7159-5p	2
hsa-miR-3679-3p	1	hsa-miR-4739	1	hsa-miR-7515	4
hsa-miR-3692-3p	1	hsa-miR-4742-3p	2*	hsa-miR-7849-3p	1
hsa-miR-3692-5p	2	hsa-miR-4747-5p	1	hsa-miR-7853-5p	3*
hsa-miR-3914	4	hsa-miR-4748	1		

Table 6.6: Predicted miRNA binding sites are enriched at the HPE-associated SNVs in the wildtype *ZIC2* 3'UTR. A two-tailed Chi-squared goodness of fit test was performed to determine whether the predicted miRNA binding sites were randomly distributed within the 3'UTR. Significance of $p<0.05$ (*) or $p<0.01$ (**) indicates predicted binding site enrichment in the region being tested. **(a)** Binding site enrichment in the non-coding conserved element (NCE; 540 bp) compared to the non-conserved region (559 bp) of the *ZIC2* 3'UTR. **(b)** Binding site enrichment in the AU-rich element (ARE; 64 bp) compared to the remaining *ZIC2* 3'UTR (1035 bp). **(c)** Binding site enrichment at the individual HPE-associated SNVs (1 bp) compared to the non-mutated 3'UTR (1089 bp). Observed and expected values are indicated. Degrees of freedom = 1 for all analyses.

A	Observed (n=420)	Expected	χ^2	*p*
NCE	218	206.37	1.289	0.2563
Non-conserved region	202	213.63		

B	Observed (n=420)	Expected	χ^2	*p*
ARE	32	24.45	2.476	0.1156
Non-ARE	388	395.55		

C	Observed (n=420)	Expected	χ^2	*p*
M1	10	0.38	243.76	0.0001**
M2	12	0.38	355.65	0.0001**
M3	6	0.38	83.19	0.0001**
M4	5	0.38	56.22	0.0001**
M5	9	0.38	195.71	0.0001**
M6	6	0.38	83.19	0.0001**
Non-mutated 3'UTR	n – mutant observed	419.62		

Table 6.7: Predicted changes in thermodynamic binding upon introduction of the six identified SNVs. Changed were predicted via mrSNP and miRNASNP databases (Deveci *et al.*, 2014; Gong *et al.*, 2015). A * indicates a change in thermodynamic strength that is unlikely to have a large impact.

Mutation	miRNA	Energy when WT	Energy When Mutated	Direction of Change
M1	hsa-miR-3163	-16 kcal/mol	-39.4 kcal/mol	Increased
M1	hsa-miR-4694	0 kcal/mol	-40.1 kcal/mol	Increased
M2	hsa-miR-3148	-41.3 kcal/mol	-17.3 kcal/mol	Decreased
M2	hsa-miR-4476	-15.29 kcal/mol	-12.42 kcal/mol	Decreased
M4	hsa-miR-3137	-15.77 kcal/mol	-15.04 kcal/mol	Decreased *
M6	hsa-miR-582	-13.9 kcal/mol	-40.8 kcal/mol	Increased
M6	hsa-miR-493-5p	-12.66 kcal/mol	-12.80 kcal/mol	Increased *

6.2.3 The *ZIC2* 3'UTR contains multiple PASs

The location of PASs in the human and mouse *ZIC2* 3'UTR were predicted using the Poly(A) Signal Miner (Liu *et al.*, 2003) and AREsite database (Gruber *et al.*, 2011) which identify putative PASs via the 'AAUAAA' and 'AUUAAA' motifs. A probability score greater than 0.6 indicates a site that is predicted to be functional. In the mouse 3'UTR, four putative PASs were annotated, of which three (mP1-mP3) are predicted to be real (Table 6.8). The second site (mP1) was not conserved between the mouse and human genomes, whilst the third and fourth sites (P2 and P3, respectively) were highly conserved between both genomes. In comparison to the mouse 3'UTR, three putative poly(A) sites were identified in the human 3'UTR and designated hP1, P2 and P3 (Figure 6.2). Whilst hP1 and P2 are likely to be real (p>0.6), P3 exhibited a low probability score and may be a sequence artefact. Nevertheless, the high level of conservation between the mouse and human P3 suggests this PAS is worth investigating further. Three additional PAS were predicted upstream of hP1 (designated hPA0.1, hPA0.2 and hPA0.3), however these were clustered together in an AU rich string of nucleotides and are unlikely to be functional (p>0.6). In summary, both human and mouse *ZIC2* 3'UTRs are predicted to contain three PASs, of which only the two most distal PASs are conserved. The 3'UTR of both species was screened for an additional nine non-canonical PASs (identified in (Beaudoing *et al.*, 2000)), however these were not predicted to occur in either human or mouse *ZIC2* 3'UTR transcripts.

hP1, mP1, P2 and P3 do not overlap directly with a predicted SNV. P2 occurs in the nine bp between M3 and M4 in both species and is also in the middle of the predicted ARE region (Figure 6.2). This, coupled with the large number of predicted RBP and miRNA binding sites, suggests that the region surrounding M3 and M4 in the *ZIC2* 3'UTR is an active site of transcript regulation and that the SNVs may affect transcript stability. Similarly, both hPA1 and mPA1 occur within 28 bp of a SNV (M1 and M2, respectively), suggesting these mutations may be able to exert an influence on alternative polyadenylation. The overall predicted effect of the six HPE-associated SNVs on *ZIC2* mRNA stability is summarised in Table 6.9.

To determine if one PAS in the *ZIC2* 3'UTR was preferentially used in mouse and human tissue, published *ZIC2* EST transcripts on UniGene (Pubmed) with known tissue of origin were assembled and their transcript length compared in multiple tissues. The sequence of each transcript was aligned to both the human and mouse *ZIC2* 3'UTR, allowing for up to five mismatches in the sequence. Upon alignment to the *ZIC2* 3'UTR, the PAS closest to the end of

Table 6.8: The mouse and human *ZIC2* 3'UTRs are predicted to contain multiple poly(A) sites. Sites were predicted using the Poly(A) Signal Miner (Liu *et al.*, 2003) and AREsite database (Gruber *et al.*, 2011). A probability score >0.6 indicates a site that is predicted to be real. Human sites are denoted by h, murine sites by m. Poly(A) sites correlate with Figure 6.2.

PAS	Start bp	PAS Type	Probability Score	Conserved between mouse and human?
hP0.1	239	AATAAA	0.499	No
hP0.2	246	AATAAA	0.505	No
hP0.3	254	AATAAA	0.520	No
hP1	810	AATAAA	0.969	No
hP2	891	ATTAAA	0.705	Yes
hP3	1062	AATAAA	0.092	Yes
mP0	85	ATTAAA	0.238	No
mP1	562	ATTAAA	0.883	No
mP2	895	ATTAAA	0.673	Yes
mP3	1065	AATAAA	0.68	Yes

Table 6.9: The six HPE-associated SNVs are predicted to influence *ZIC2* transcript stability. SNVs that are predicted to alter the destabilising or stabilising influence conferred by each element onto the *ZIC2* mRNA are denoted with an X.

Stability element	M1	M2	M3	M4	M5	M6
ARE region			X	X		
RBPs	X	X	X	X	X	X
miRNA	X	X		X		X
Poly(A) sites	X	X	X	X		

the transcript was determined and correlated to the tissue of transcript origin. A total of thirteen adult samples (but no embryonic samples) were available for the human *ZIC2* transcript (Table 6.10). All five brain tissue samples and all three bone tissue samples utilised the most distal PAS (P3), whilst only one of five lung samples terminated at this site (Table 6.10). The remaining four lung transcripts appear to terminate midway through the 3'UTR where no PAS is predicted, suggesting that there may be a yet unidentified PAS in this region or an experimental artefact. No other tissue types were able to be aligned to the *ZIC2* 3'UTR, and only one mouse record was available, preventing analysis of mouse PASs.

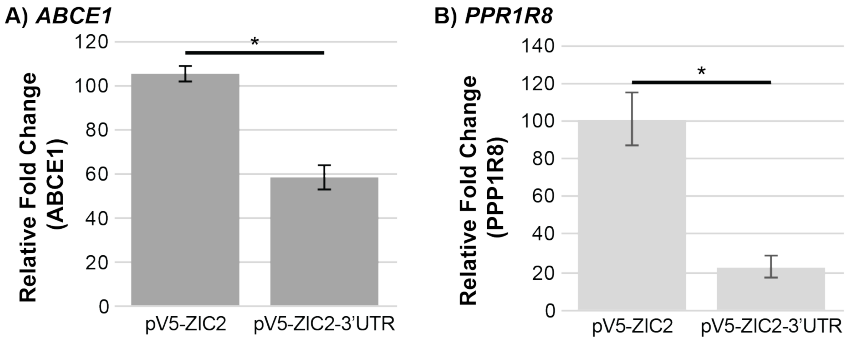
6.2.4 The 3'UTR Influence *ZIC2* Transcript Decay Rates

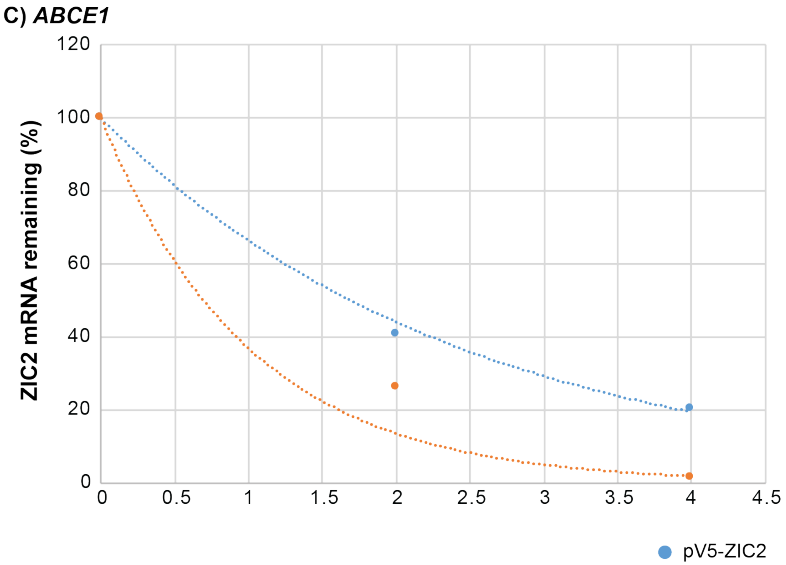
To analyse whether the 3'UTR is involved in post-transcriptional regulation of *ZIC2* transcripts, an assay was designed to evaluate transcript levels following transcriptional blockade in cultured mammalian cells and relative quantification of *ZIC2* transcript. It was expected that the presence of the 3'UTR, which contains the putative ARE region, will alter steady-state levels of *ZIC2* transcript and significantly increase the rate of decay. Optimization of the RT-qPCR protocol, selection of appropriate reference genes and oligonucleotide efficiency testing were all performed in conjunction with Kathryn Dickson (Appendix Section 4.1-4.3). Total RNA was extracted from HEK293T cells transfected with pV5-*ZIC2* or pV5-*ZIC2*-3'UTR (Appendix Figure A4.3) and checked for integrity and lack of gDNA contamination prior to RT-qPCR analysis. Cells transfected with pV5-ZIC2 were exhibited a significantly higher fold change in *ZIC2* expression above background levels than cells transfected with pV5-*ZIC2*-3'UTR (Figure 6.3 A and B), indicating that the presence of the 3'UTR decreased steady-state *ZIC2* transcript levels.

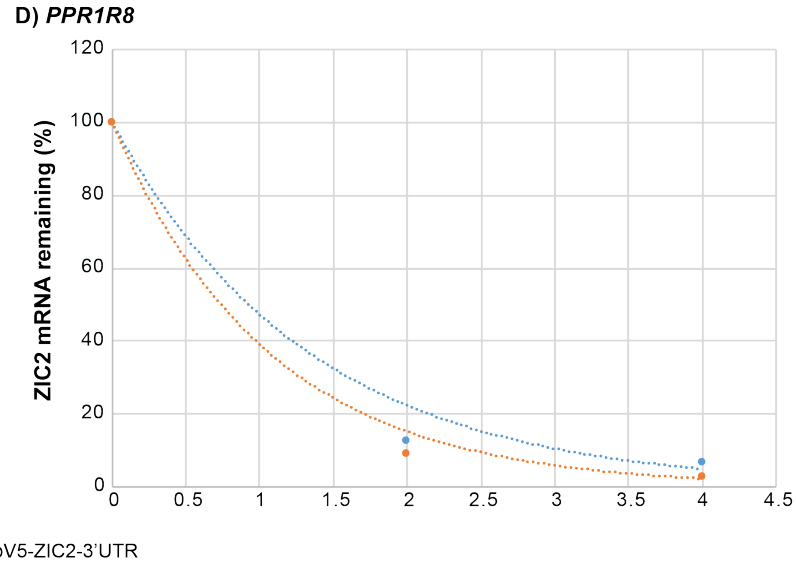
To determine if this difference was due to transcript degradation, transcription was halted via the addition of the transcriptional inhibitor DRB (5,6-dichloro-1-beta-D-ribofuranosylbenzimidazole) and mRNA decay rates assessed. HEK293T cells transfected with pV5-*ZIC2* and pV5-*ZIC2*-3'UTR, along with untransfected cells, were treated with 200 µM DRB or the vehicle control (0.2% DMSO) 24 hr post transfection and RT-qPCR used to quantify the fold change in *ZIC2* expression at multiple time points (0 hours [not treated with DRB], 2 and 4 hours). The level of *ZIC2* transcript in DRB treated cells was calculated as a percentage of the level in vehicle only treated cells from the equivalent time point, and plotted over time. The fold changes of *ZIC2* expression quantified in cells treated with fresh DRB were lower than those in the equivalent vehicle control treated cells, indicating DRB was having an inhibitory

Table 6.10 Human EST clones and tissue expression.

GenBank ID	cDNA Clone	End Position	Poly(A) site	Tissue	Nearby
AI458310.1	IMAGE:2150307	414	-	Lung	41 bp upstream (5') of M1
AA885663.1	IMAGE:1500176	482	-	Lung	26 bp downstream (3') of M1 95 bp upstream (5') of M2
AA900219.1	IMAGE:1604674	588	-	Lung	10 bp downstream (3') of M2
BE219194.1	IMAGE:3176777	624	-	Lung	46 bp downstream (3') of M2
BF594212.1	IMAGE:3564059	1075	PA3	Brain	3' transcript end at 1082
BF438724.1	IMAGE:3275267	1081	PA3	Brain	3' transcript end at 1082
BU619638.1	UI-H-FH1-bfq-o-01-0-UI	1082	PA3	Bone	3' transcript end at 1082
CA427434.1	UI-H-FH1-bfi-b-11-0-UI	1082	PA3	Bone	3' transcript end at 1082
AA491603.1	IMAGE:910539	1082	PA3	Bone	3' transcript end at 1082
R42515.1	IMAGE:29730	1082	PA3	Brain	3' transcript end at 1082
AI359104.1	IMAGE:2012588	1082	PA3	Brain	3' transcript end at 1082
R61372.1	IMAGE: 37980	1082	PA3	Brain	3' transcript end at 1082
BG149342.1	IMAGE:3366742	1082	PA3	Lung	3' transcript end at 1082







E) Half-life

	Construct	pV5-ZIC2	pV5-ZIC2-3'UTR
Reference Gene	ABCE1	1.731 hr	0.641 hr
	PPP1R8	1.025 hr	0.791 hr

Figure 6.3: The 3'UTR decrease *ZIC2* mRNA half-life. (a-b) ZIC2 expression was normalized to **(a)** *ABCE1* or **(b)** *PPP1R8*. The fold change was calculated via $2^{-\Delta\Delta CT}$. **(c-d)** The percentage of *ZIC2* mRNA remaining over time after treatment with DRB, calculated relative to *ZIC2* expression in DMSO treated cells. DRB (200 μM) or DMSO only was used to treat HEK293T cells, 24 hours post transfection with pV5-*ZIC2* (blue line) or pV5-*ZIC2*-3'UTR (orange line). RNA was extracted from cells at 0, 2 and 4 hours following DRB treatment and the fold change in *ZIC2* expression quantified by RT-qPCR at each time point as $2^{-\Delta\Delta}$ CT. The fold change in *ZIC2* expression in DRB treated cells was calculated as a percentage of that quantified from DMSO only treated cells transfected with the same base construct, at the equivalent time point. The trend line shows the exponential relationship of the proportion of *ZIC2* mRNA remaining over time. **(c)** The percentage of *ZIC2* mRNA remaining normalised using *ABCE1*. **(d)** The percentage of *ZIC2* mRNA remaining normalised using *PPP1R8*. **(e)** The half-life of *ZIC2* mRNA in the presence and absence of the 3'UTR. N=3. *: $p<0.001$, Student's T-Test; n=3; error bars: S.E.M.

effect on transcription as expected (not shown). The percentage of *ZIC2* mRNA remaining in inhibited cells transfected with either construct appeared to decrease in an exponential fashion over time (Figure 6.3 C and D), therefore the half-life of *ZIC2* mRNA in cells transfected with the constructs was approximated from the nonlinear regression of the data points. *ZIC2* mRNA in cells transfected with pV5-*ZIC2*-3'UTR was found to have a shorter half-life than in cells transfected with pV5-*ZIC2* (Figure 6.3 E), indicating that the inclusion of the 3'UTR leads to increased *ZIC2* mRNA degradation within a cell. As constructs containing the mutated 3'UTR were unable to be created, analysis of how the six SNVs influenced transcript decay rates could not be performed.

6.2.5 The *Zic2* 3'UTR employs the most distal poly(A) site at gastrulation

To determine which of the three mouse PASs were most likely to be used during embryo development, two approaches were taken: Rapid Amplification of cDNA Ends (RACE) and relative quantification of the *Zic2* transcript at multiple locations within the 3'UTR via RT-qPCR. For RACE, the *Zic2* 3'UTR was PCR amplified from wildtype C3H/HeH murine embryo mRNA (at 7.5 dpc, 8.5 dpc and 9.5 dpc) that had been converted to cDNA using a standard RACE dT-adaptor oligonucleotide, and Sanger sequenced. Though the sequences obtained did initially align to the *Zic2* 3'UTR, accurate full-length sequences were unable to be obtained due polymerase slippage at the 3' end and no determination could be made regarding the poly(A) that was being utilised.

As an alternative method, four pairs of oligonucleotides were designed to amplify each region of the 3'UTR 5' to the predicted poly(A) site via RT-qPCR (Appendix Table A1.2, Figure 6.4). It was hypothesised that if the most distal poly(A) site was being used (mP3), then all four pairs of oligonucleotides would result in amplification of the same relative amount of *Zic2* mRNA. Conversely, if the first poly(A) site was the predominant poly(A) site active at gastrulation (P1), it was expected that the levels of *Zic2* mRNA at P0 and P1 would be high in contrast to low *Zic2* levels at P2 and P3 sites. RT-qPCR was performed on total mRNA extracted from embryos at 6.5, 7.5, 8.5 and 9.5 dpc, and the results normalised to *H2afz* or *Ubc* via the $2^{-\Delta\Delta CT}$ method. *H2afz* and *Ubc* were chosen as reference genes as they are stably expressed across multiple time points at gastrulation (Barratt et. al. unpublished manuscript, Appendix Section A4.4). Fold changes in mRNA levels were calculated relative to the lowest expression value. Upon analysis, a trend can be seen across the four age groups analysed where the level of *Zic2* mRNA proximal to P3 is equal to or increased (as can be seen at 8.5 dpc) compared to *Zic2* mRNA levels at P0 and P1, suggesting that the most distal poly(A) is the site being utilised across all four ages of embryo development (Figure 6.4). A significant peak in mRNA expression at P2

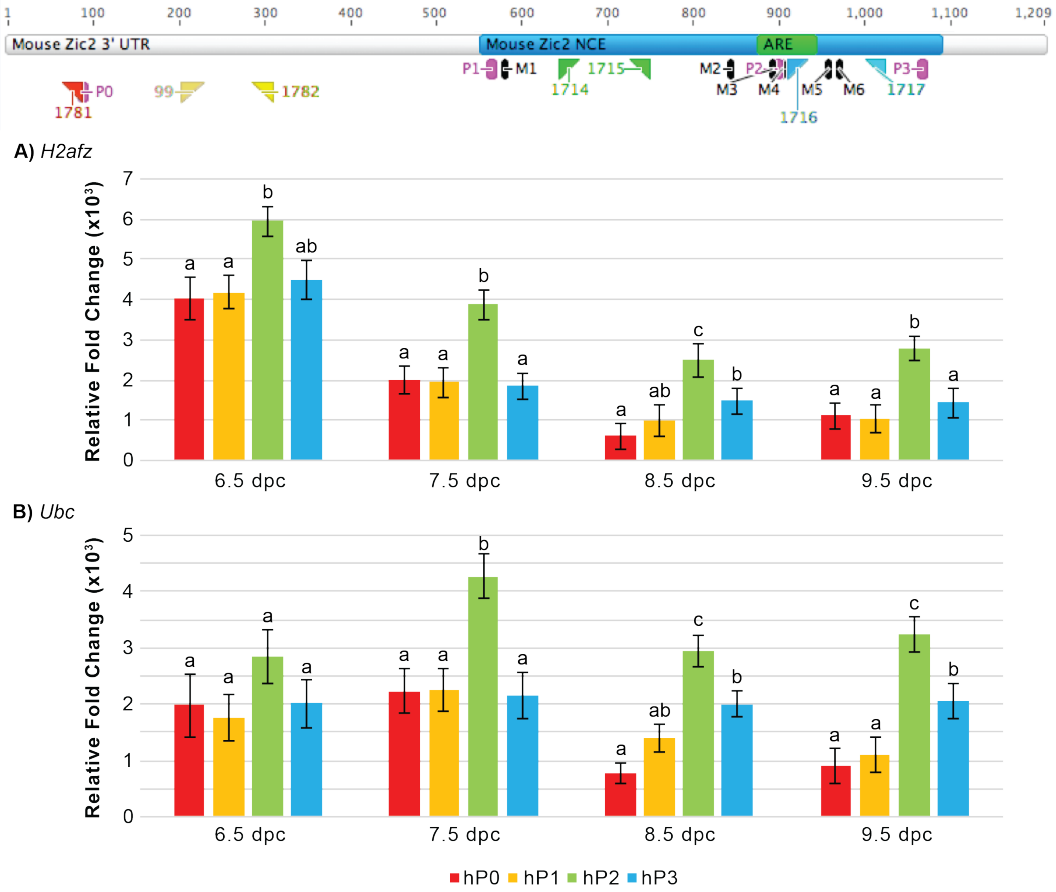


Figure 6.4: The distal *Zic2* 3'UTR poly(A) site P3 is the most likely site used during gastrulation. Total mRNA was extracted from murine embryos at the stages show and *ZIC2* 3'UTR mRNA levels upstream of each predicted PAS measured via RT-qPCR. Results were normalised to *H2afz* or *Ubc* via the $2^{-\Delta\Delta CT}$ method, and fold change calculated relative to the lowest expression value. Error bars = S.E.M of three external repeats; letters denote statistical significance of $p<0.05$ calculated via a one-way ANOVA with Fischer's unprotected post ad hoc test. The letter a denotes statistical significance to b and so on within each age group (e.g. 6.5 dpc) but not between age groups (ie. the letter a at 6.5 dpc does not correlate to a at 7.5 dpc, etc). Colours of forward and reverse oligonucleotides (triangles) on the 3'UTR correlate with hP0, hP1, hP2 and hP3 in (a) and (b).

was detected from 7.5 - 9.5 dpc. Whilst a peak such as this may indicate non-specific mRNA amplification at off-target sites or primer dimers (investigated in Appendix Section A4.5 and Table A4.1), it may also indicate that additional non-coding mRNA is being produced at this region of the *Zic2* 3'UTR, such as eRNA.

6.2.6 Long non-coding RNA transcripts correlate with the *Zic2* 3'UTR

To investigate whether the peak seen at P2 correlates with predicted transcription of eRNA or other non-coding RNAs, the UCSC Genome Browser was utilised. As no tracks which annotate Long RNA-seq contigs in the mouse *Zic2* 3'UTR were able to be identified, the CSHL Long RNA-seq (>200 nt) and CSHL Small RNA-seq (<200 nt) tracks were used with the human *ZIC2* 3'UTR (assembly GRCh37/hg19) (Cheng *et al.*, 2015; Langmead *et al.*, 2009; Parkhomchuk *et al.*, 2009). Both tracks contain data from a range of cell types, including but not limited to H1-hESCs, GM12878 and HeLa cell lines. RNA-seq contigs with their 5' in the CDS of *ZIC2* were excluded from analysis, as eRNAs by definition must occur within an ER element. Upon analysis, eight differently sized long RNA and six differently sized small RNA molecules were predicted to be transcribed from within the *ZIC2* 3'UTR (Figure 6.2 C). When the human sequence was aligned to the mouse 3'UTR, the majority of the long RNA molecules overlap with the P2 oligonucleotide amplicon position in both the human and mouse *Zic2* 3'UTR, suggesting that the significant peak detected at P2 in the RT-qPCR experiment may indeed be from eRNA or a similar non-coding RNA (this assumes the production of eRNA is conserved between man and mouse). Interestingly, one predicted long RNA molecule overlaps with the P3 oligonucleotides in both the human and mouse *Zic2* 3'UTR, which may account for the small but significant increase in *Zic2* transcript detected at P3 in 8.5 dpc embryos.

6.3 Discussion

During embryo development, mRNA transcripts and proteins are required in a spatially and temporally restricted manner, with errors in this expression often leading to catastrophic defects. To prevent such errors, an intricate system of mRNA decay mechanisms has evolved. In this chapter, I have shown that the *ZIC2* 3'UTR promotes *ZIC2* transcript decay, most likely via a combination of AREBPs and miRNAs interacting with an ARE within the 3'UTR. Additionally, I have shown that the distal PAS of the 3'UTR is used during murine gastrulation, and that an increase in transcript levels within the 3'UTR correlates to a predicted long non-coding RNA that is likely to be eRNA.

6.3.1 The *Zic2* 3'UTR promotes transcript decay

The presence of the *ZIC2* 3'UTR correlates with a decrease in *ZIC2* mRNA half-life (0.64 hr compared to 1.73 hr when normalised to *ABCE1*). This agrees with multiple studies which found a negative correlation with increased 3'UTR length and mRNA stability (Michalova *et al.*, 2013; Sharova *et al.*, 2009). The removal of the 3'UTR from the *ZIC2* transcript likely prevents trans-elements involved in 3'UTR mediated destabilisation such as miRNAs and RBPs from targeting the element and destabilising the transcript. The half-life of *ZIC2* transcripts when in the presence of the 3'UTR differs from the rate of 1.97 hours previously determined via a microarray based study using mouse embryonic stem cells (Sharova *et al.*, 2009). As it is known that tissue specific factors are involved in mediating mRNA decay (Barreau *et al.*, 2005), the half-life of *ZIC2* in various cell types cannot be expected to be similar. Furthermore, as Sharova et al examined endogenous expression they did not assess whether the presence of the 3'UTR influenced the stability of the *ZIC2* transcript. The results presented in this chapter do agree with one conclusion from Sharova et al, however, which is that *ZIC2* exhibits a relatively short half-life (<2 hr) compared to the median half-life for mRNA turnover (~10 hrs) in cultured cells (Sharova *et al.*, 2009; Yang *et al.*, 2003). A short half-life is expected for TFs and developmental genes, as it allows rapid up- and down-regulation in response to environmental triggers (Yang *et al.*, 2003).

Whilst not addressed in this study, it is presumed that the reduction in *ZIC2* transcript half-life seen in the presence of the 3'UTR can be attributed in part to the presence of the ARE. As such, an investigation into whether the putative ARE outlined above is functional and the minimum sequence required to promote mRNA decay should be conducted. By deleting the surrounding regions of the UTR that are not predicted to be part of the ARE from the pV5-*ZIC2*-3'UTR constructs, or deleting the ARE region from the 3'UTR, the effect the putative ARE region has on *ZIC2* transcript decay can be determined. If this ARE is indeed the main driver of *ZIC2* transcript decay then it is likely that deletion of this fragment will produce a significant increase in *ZIC2* mRNA half-life. In comparison, if the surrounding regions are not essential for transcript decay

rates then the removal of these regions should produce a decay similar to what is seen with the entire 3'UTR still present. An alternative form of this same experiment is to clone the predicted ARE region of the *Zic2* 3'UTR into the rabbit *β-globin* gene, which has a strictly defined half-life and is commonly used to analyse ARE domains (Chen *et al.*, 1994; Li *et al.*, 2013; Ysla *et al.*, 2008). If the half-life of the *β-globin* transcript is dramatically altered by the introduction of this ARE region, then this will confirm that it is indeed a true ARE and the driving force behind ZIC2 transcript decay.

Not all of the PASs predicted in the 3'UTR appear to be active during the early stages of embryogenesis. The distal 3' PAS is considered the default site for most transcripts due to the availability of regulatory sites within the full length 3'UTR (Beaudoing *et al.*, 2000; Weill *et al.*, 2012). A preference for the distal PAS at gastrulation agrees with the idea that *Zic2* is a transiently expressed TF that must undergo carefully controlled rapid expression followed by rapid decay to pattern the embryo. Whether P3 remains the default PAS at other embryonic or adult stages remains to be elucidated, however. Single-Molecule sequencing chemistry with Real-Time detection (SMRT; PacBio) and Full-length Isoform Sequencing (Iso-Seq; PacBio) can be employed to determine the precise PAS and cleavage sites that are utilised in both human cell lines and murine embryos. These techniques allow the sequencing of long stretches of alternatively transcribed RNAs without the requirement of PCR amplification and are suitable for genes with a high GC content and large amounts of repetitive sequences such as the ZICs. Briefly, poly(A)$^+$ RNA molecules are extracted from the desired tissues and converted to cDNAs via priming with Anchored Oligo(dT)$_{20}$ oligonucleotides. A SMRTbell sequencing library is generated, with templates complexed to polymerases and bound to magnetic beads. DNA sequencing is then carried out with a PacBio RS II sequencer, producing transcriptome data that can then be analysed. This technique is optimised to pick up any polyadenylated lncRNA molecules (Tombácz *et al.*, 2016) produced from the *ZIC2* 3'UTR and can presumably be employed on total RNA extracts for detection of bidirectional un-polyadenylated lncRNAs such as eRNA. Additionally, the sensitivity of the techniques will allow for the analysis of finely dissected embryonic tissues with relatively few cells, such as the node. Identification of the predominant transcript in the node will aid in narrowing down regions of the *Zic2* 3'UTR where mutations will most likely effect node development and function. Finally, if CRISPR can be utilised to introduce the six HPE-associated SNVs into human cell lines or mouse embryos, the influence of each mutation on PAS and transcript quantity can be measured. Once the poly(A) cleavage sites in use in the *ZIC2* 3'UTR have been elucidated, the effect of transcript isoforms on mRNA decay rate can be determined by PCR amplifying the 3'UTR truncated at each PAS, cloning it into a plasmid containing *Zic2* cDNA, and analysing transcript stability via the decay assay

established in this chapter. It is expected that the shorter the transcript, the longer the half-life as there is a reduction in the number of available miRNA and RBP binding sites.

Due to cloning difficulties the influence of the six HPE-associated SNVs on the *ZIC2* decay rate could not be established. This goal therefore remains a paramount objective before further decay analysis is attempted. Interestingly, previous analysis on *c-fos* AREs found that specific nucleotide changes had a dramatic effect on transcript stability. Mutations in which an A or U are changed to a G or C dramatically increase the stabilization of a transcript (Chen *et al.*, 1994), presumably due to a reduction in the AU richness and RBP/miRNA binding sites that define ARE domains. As such, it is expected that M3 (U>C) and M4 (A>G) would exert the largest effect on stability as they are located within the predicted ARE region, potentially increasing the half-life of *ZIC2* transcripts and resulting in an increase in ZIC2 protein. All six SNVs, however, have the potential to alter transcript decay, whether it be by altering the RBPs and miRNAs predicted to bind the 3'UTR, or by altering the PAS used. Whilst none of the mutations occur directly within a predicted PAS, it is likely that those mutations that occur within the 15-30 bp stretch between a PAS and cleavage site (such as M4) could presumably alter polyadenylation either by increasing transcript length or increasing transcript stability. For example, Orkin et al. previously identified a T>C substitution in a conserved PAS in the *β-globin* 3'UTR that sufficiently disrupts polyadenylation, resulting in a 900 bp increase in transcript length (Orkin *et al.*, 1985). Alternatively, it is possible that the mutations near the distal PAS will result in a shift to a 5' PAS. Whilst the mutant transcript produced would still undergo decay, its half-life would most likely increase due to the loss of miRNA and RBP binding sites. As such, the transcript would exist in the cells for longer than necessary, resulting in an increase in protein above the threshold required. To circumvent the cloning difficulties encountered in this chapter, the mutant V5-ZIC2-3'UTR constructs required for the decay assay would need to be bought from a commercial company, or CRISPR-Cas9 used to mutate HEK293T cells directly.

6.3.2 Multiple RBPs and miRNAs are predicted to interact with the *ZIC2* 3'UTR

Bioinformatic analysis identified 40 RBPs and 248 miRNAs predicted to bind the *ZIC2* 3'UTR. It should be noted, however, that these are mere predictions and not indicative of in vitro or in vivo binding. As one functional TFBS is estimated to occur per 1000 predicted sites (Wasserman and Sandelin, 2004), it is presumed that the rate of true RBP binding sites is similar. Nevertheless, to determine the exact mechanisms behind targeted decay of *ZIC2* during embryogenesis, the candidate RBPs and miRNAs that are predicted to bind the *Zic2* 3'UTR and influence its transcript decay need to be narrowed down. HEK293 cells lack TTP family members, therefore the candidate pool of RBPs bound to the *ZIC2* 3'UTR is already reduced in this cell line (Al-Souhibani *et al.*, 2010; Blackshear, 2002). Pulldown of biotinylated RNA oligonucleotides that

correspond to the *ZIC2* ARE followed by western blot analysis with antibodies for HEK293 specific RBPs such as AUF1 can be used to determine whether multiple isoforms of each RBP are bound to the *ZIC2* ARE, and whether preferential binding between isoforms is occurring (based on the predicted molecular weights of the bands observed on band intensities, respectively), as seen in experiments of the TSP-1 3'UTR (McGray *et al.*, 2011). Similarly, biotinylated miRNAs can be used to pulldown proteins they bind to, with a ZIC2 antibody used in western blot analysis to determine if the miRNA is binding the 3'UTR. Two caveats exist with this method, however: (i) it requires a ZIC2 antibody, of which a satisfactory commercial antibody is currently lacking, and (ii) the narrowing down of candidate miRNAs from 248 to a number that can be reasonably analysed. Once candidate RBPs and miRNAs have been sufficiently narrowed down, their effect on ZIC2 transcript half-life can be determined via introduction into the decay assay.

6.3.3 Are *ZIC3* and *ZIC5* transcripts subject to decay?

There is strong evidence for overlapping function between the *ZIC* family, particularly between *ZIC2*, *ZIC3* and *ZIC5* during gastrulation. Additionally, the Muenke laboratory (NHGRI, NIH) have identified a putative NCE region within the *ZIC3* 3'UTR that contains Heterotaxy-associated SNVs (Roessler and Muenke, personal communication). Comparatively, no mutations in *ZIC5* have been linked to human hydrocephaly despite homozygous *Zic5* cDNA mutations in the mouse resulting in hydrocephaly (Arkell lab, unpublished data). It is therefore unlikely that the non-coding regions of *ZIC5* have been analysed in human hydrocephaly patients. Consequently, it is possible that mutations in a regulatory region of the *ZIC5* 3'UTR are contributing to human cases of this disease and remain unidentified. If the 3'UTRs of *ZIC3* and *ZIC5* are found to significantly contribute to transcript decay via the mechanisms identified above, it would provide direction for future mutagenesis screens in human patients.

6.3.4 Conclusion

Post-transcriptional regulation of mRNA is used by cells to strictly control the spatial and temporal expression of genes. Via transcript decay mechanisms, gene expression can be rapidly turned off when the protein is no longer required. Defects in this process, however, can lead to exogenous gene expression or levels of protein above the required threshold, resulting in disease. From the analysis shown in this chapter, it appears that *ZIC2* expression and transcript half-life is regulated by multiple mRNA stability elements residing within the 3'UTR including AREs, alternative poly(A) sites and eRNAs. The use of the most 3' PAS during gastrulation produces a transcript that contains all of the predicted stability elements, correlating with a need for rapid degradation of the *Zic2* transcript during this time to restrict gene expression to specific tissues of the embryo.

Chapter 7: In vivo analysis of the *Zic2* NCE

The work in this chapter on *PiggyBac* transgenics was performed with the Muenke laboratory at the National Human Genome Research Institute (NHGRI) and the Transgenic Core Facility (NIH, Maryland, USA), whilst the CRISPR mutagenesis work was performed in with the Burgio laboratory at the John Curtin School of Medical Research (JCSMR) Transgenesis Facility (ANU, Canberra, Australia). These facilities aided in the production of transgenic and CRISPR mutants.

7.1 Introduction

To establish if the *ZIC2* NCE has enhancer activity, Roessler et al. (2012a) previously created transient transgenic zebrafish by cloning the putative NCE region into a ZED vector containing both GFP (reporter of *ZIC2* NCE activity) and RFP (internal control). The wildtype *ZIC2* NCE was inserted into the vector in both the forward (5' – 3') and the reverse orientation (3' – 5') and transgenic zebrafish embryos analysed. NCEs in the forward orientation that contained the six individual HPE-associated SNVs were also inserted. Despite a high level of transgenesis, GFP expression specific for the *ZIC2* NCE could not be reliably detected in the resulting embryos. Further attempts to detect low level GFP signal via WMISH resulted in artificial signal from the vector backbone (Roessler *et al.*, 2012a) or inconsistent expression between founder lines (SK. Hong, E. Roessler and M. Muenke, personal communication, Appendix Figure A5.1). The results suggested that (i) the ZED constructs are not the best method of transgenesis to use in such experiments, (ii) zebrafish are not the best animal to model *ZIC2* NCE activity, or (iii) the *ZIC2* NCE does not regulate *ZIC2* expression during embryogenesis. In parallel, experiments to test the construct in mice were planned. Traditional transgenic and mutagenesis approaches in mice exhibit relatively low success rates, and can take 9-12 months before viable offspring are available for analysis. Over the past decade, however, new genome editing technologies have emerged as useful tools for high-throughput and specific editing of the mouse genome. I attempted to use two different mouse mutagenesis strategies to evaluate *ZIC2* NCE function and the pathogenesis of the six SNVs in vivo: *PiggyBac* transgenics (gain-of-function) and CRISPR-mutagenesis (loss-of-function).

7.1.1 The *PiggyBac* transposase

A transposon is a sequence of DNA or RNA with the ability to change its position in the genome. A transposon consists of two elements – a DNA transgene consisting of the desired cargo sequence flanked on either side by two inverted terminal repeats (ITRs), and a transposase that catalyses the insertion of the DNA component into the target genome via a genetic 'cut and paste' method (Figure 7.1). This system can be modified for transgenesis through the co-

injection of the synthesized transposase mRNA with the circular transgene DNA (the cargo component and ITRs) into the pronuclei of zygotes of the preferred model organism (Mátés, 2011). Genome integration is essentially random, although *PiggyBac* does require a TTAA nucleotide sequence in order to correctly insert its DNA cargo into the genome (Bjork *et al.*, 2010; Fraser *et al.*, 1996).

Transposons are commonly used for transgenesis and insertional mutagenesis in a wide range of model organisms (e.g. *Drosophila, Caenorhabditis elegans* and plants). Due to a lack of naturally active transposons in mammals, however it is only recently that they have begun to be utilized in transgenic mice and rats (Dupuy *et al.*, 2002; Mátés, 2011). Two transposases previously used to generate transgenic mice are *SleepingBeauty* and *PiggyBac*. *SleepingBeauty* was reconstructed over two decades ago from defective transposon sequences found in the salmonid genome (Ivics *et al.*, 1997). Although it is the most common transposon system used in mammalian experiments, *SleepingBeauty* offers a germ-line transgenic efficiency of only 15% (Dupuy *et al.*, 2002). Since its discovery, a hyperactive form has been developed which increased transgenic efficiency 100-fold. Though the hyperactive form works successfully in cultured cells, the efficiency often fails to translate into stable gene transfer of primary cells *in vivo* (Mátés, 2011; Mátés *et al.*, 2009). In contrast, the *PiggyBac* transposase, isolated from the cabbage looper moth (*Trichoplusia ni*) is unique due to its high efficiency rate, large cargo capacity, and ability to form functional protein fusions and restore a donor site to its original condition upon excision, leaving no trace of transposon insertion (Cadiñanos and Bradley, 2007; Ding *et al.*, 2005; Wang *et al.*, 2008; Wu *et al.*, 2006). The *PiggyBac* transposase has a reported transgenic efficiency of up to 65% depending on the DNA component being inserted (Ding *et al.*, 2005), and has successfully demonstrated long term integration into the murine genome (Doherty *et al.*, 2012).

Transposons have greatly reduced the production time required to generate mutated mammalian germlines (Horie *et al.*, 2003), and have allowed for the production of forward genetic screens in vivo to identify genes involved in diseases such as solid tumours (Collier *et al.*, 2005). Transgene genomic integration is often subject to position effects, however, where endogenous regulatory elements act on the promoter of an inserted transgene to produce

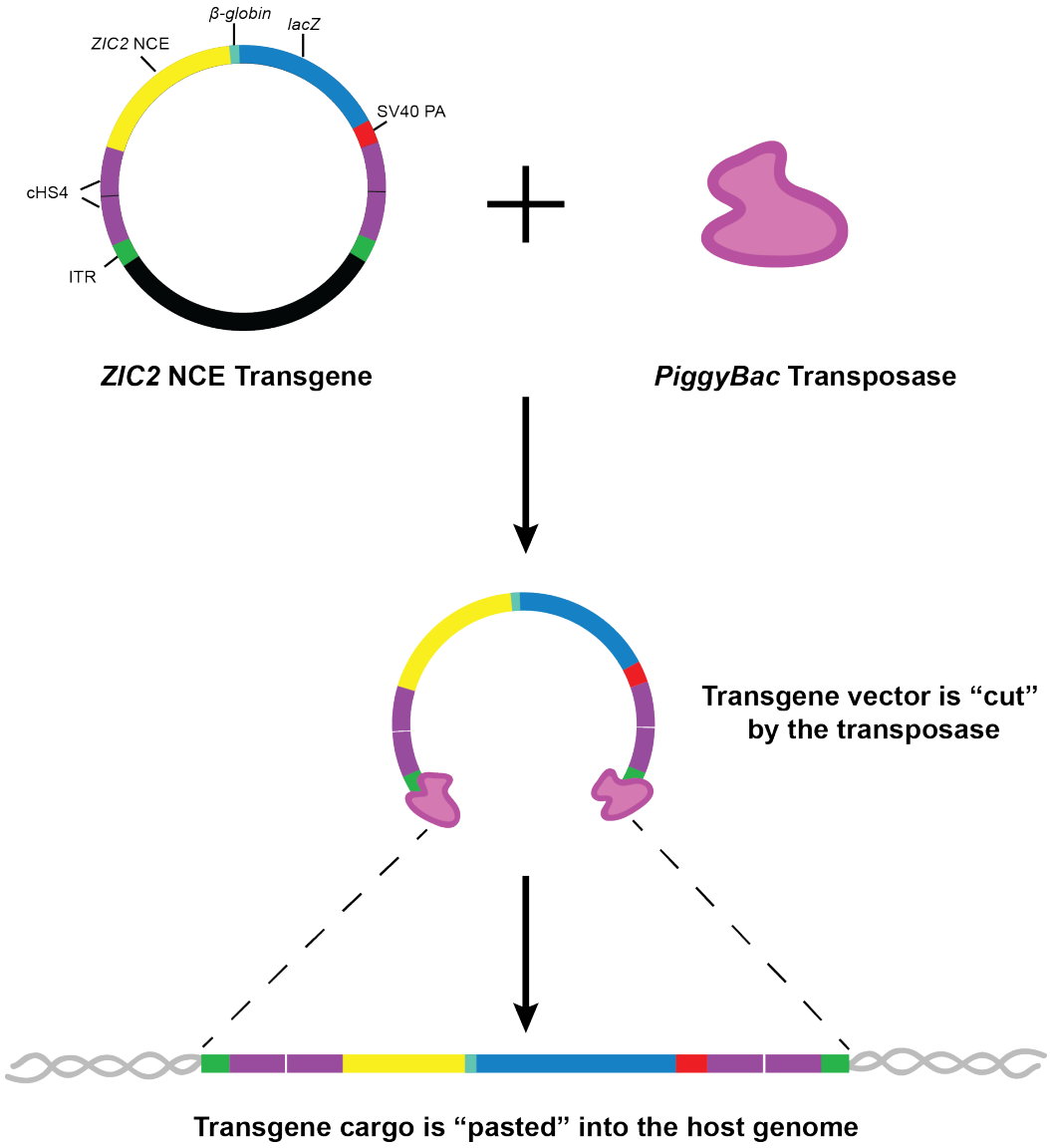


Figure 7.1: *PiggyBac* uses a "cut and paste" mechanism. The *PiggyBac* transposase recognises and binds to transposon specific ITRs on the transgene vector. It then excises the cargo, in this case *ZIC2* NCE-*lacZ* flanked by cHS4 sites, from the vector and integrates the cargo into the host genome at TTAA chromosomal sites. ITR: inverted terminal repeats; cHS4: Chicken insulator sites; *β-globin*: *β-globin* minimal promoter; SV40 PA: SV40 PolyA fragment.

false reporter expression. To mitigate this effect in the transposon transgenic system, insertion of insulator sequences that flank the cargo component of the transgene vector can be utilised (Dupuy *et al.*, 2002). One such insulator, the chicken hypersensitive site-4 (cHS4) chromatin insulator, is thought to be a good candidate to mitigate these effects (Aker *et al.*, 2007; Bjork *et al.*, 2010), with previous incorporation of the insulator into *PiggyBac* vectors resulting in a 1.5 fold increase in transgene expression in human cell lines (Sharma *et al.*, 2012).

7.1.1.1 Gain-of-function transgenic approach – the ZIC2 NCE and PiggyBac transposons

Due to the increased efficiency of the *PiggyBac* transposon, I aimed to (i) establish a high efficiency system for the creation of mouse transgenics and (ii) utilise this method to determine whether the *Zic2* NCE regulates *Zic2* expression during gastrulation, and if the six HPE-associated SNVs disrupt this regulation. Transposon cargo constructs containing a transgene made up of the *Zic2* NCE followed by a *β-globin* minimal promoter and *lacZ* reporter were created (Figure 7.1). Four cHS4 insulator fragments were included in the construct to minimize position effects based on data from Bryan Bjork (Bjork et al. 2010; B. Bjork, personal communication) who found that the inclusion of two cHS4 fragments aided in the production of *PiggyBac* transgenic embryos.

7.1.2 CRISPR-Cas9 mutagenesis

Via CRISPR-Cas9 mutagenesis, DNA can be manipulated to create gene knockouts or used to introduce foreign DNA (such as reporters), deletions and point mutations into a specific region of the genome. The CRISPR-Cas9 system uses RNA-guided nucleases to cleave specific genetic elements and allows for the mutation, insertion and removal of components of the target organisms genome. Originally isolated from the bacterial genome, the CRISPR Associated protein 9 (Cas9) is guided by trans-activating CRISPR RNA (tracrRNA) that provides a stem-loop structure that the Cas9 protein binds to, and CRISPR RNA (crRNA) containing the target sequence and tracrRNA recognition sequence. When the tracrRNA and crRNA hybridize, they form a guide RNA that allows the Cas9 protein to target and degrade foreign nucleic acids (Gasiunas *et al.*, 2012; Jinek *et al.*, 2012). This system has been modified by fusing tracrRNA and crRNAs to form a single-guide RNA (sgRNA) that can be designed to induce targeted single or double stranded breaks (DSBs) in the genome via Cas9 (Jinek *et al.*, 2012). Upon breaking, the cells endogenous non-homologous end joining (NHEJ) repair pathway ligates the DNA strands back together. While this may result in precise end joining or targeted deletions, it can also lead to the insertion of point mutations, disrupting the target's reading frame. As such, it is an effective method to use for the creation of gene knockouts (Singh *et al.*, 2014). The DSBs can also be repaired via manipulation of the cells endogenous homology-directed repair (HDR) mechanisms. By introducing an alternative repair template with locus-specific homology alongside the Cas9

protein, exact point mutations and indels can be introduced in the hosts genome. To target Cas9 to a specific DNA sequence, however, the target sequence must be immediately 5' to a 3 bp protospacer adjacent motif (PAM) sequence (defined as NGG) and be complementary to the sgRNA sequence (Yang *et al.*, 2014).

Whilst extremely versatile, CRISPR-aided mutagenesis can also be unpredictable, with superfluous sequence changes routinely found at the site of repair due to the error prone NHEJ repair system. Furthermore, founder animals are frequently mosaic for the introduced mutation (Jacobi *et al.*, 2017; Mianné *et al.*, 2017). Off-target effects, which occur when the guide sequences have sufficient homology to sites in the genome that are not the target, are also common, with overexpression of Cas9 and guide RNA correlating with a high incidence of these events (Fu *et al.*, 2013). Despite these downsides, CRISPR-mutagenesis boasts an efficiency rate of up to 100% for both HDR and NHEJ in mice, dependant on the mode of delivery and method of editing employed (reviewed in Singh et al. 2014).

7.1.2.1 Loss-of-function transgenic approach – Deletion of the Zic2 NCE via CRISPR-Cas9

Due to the versatility and control provided by CRISPR-mutagenesis, I aimed to utilise this technique to delete the *Zic2* NCE from the mouse genome and determine what role, if any, it plays in HPE development. To do this, both direct deletion and tiling CRISPR deletion approaches were employed, whereby a set of target guides spanning the length of the murine *Zic2* 3'UTR were designed and introduced into zygotes together, along with Cas9 protein, via microinjection or electroporation. The resulting DSBs would result in a series of different deletions within the 3'UTR, include the entire NCE, which would allow us to map the subsequent phenotypes and gene expressions patterns to a particular region within the 3'UTR (Figure 7.2). The tiling CRISPR approach has successfully been used to target NCEs in the past, although to a much larger degree, such as the creation of 1176 distinct deletions of the *CXCL1* 3'UTR in BEAS-2B cells (Zhao *et al.*, 2017).

7.1.3 Chapter 7 aims

As previous attempts to model the activity of the *ZIC2* 3'UTR in zebrafish have failed, the challenge is to now produce mouse models that accurately report the target gene being regulated by the *ZIC2* 3'UTR and any effects that occur when the 3'UTR, and specifically the NCE, is altered.

The aims of this chapter were:

- To generate a highly efficient system for the creation of mouse transgenics via the *PiggyBac* transposase.

- To use *lacZ* transposon transgenics to determine if the *ZIC2* NCE drives expression of *Zic2* during gastrulation and whether the six mutated forms of the *ZIC2* NCE alter *lacZ* reporter expression from the wildtype NCE.
- To optimize the deletion of non-coding regions in mice via CRISPR-Cas9.
- To determine what effect deletion of the endogenous *Zic2* NCE has on embryo development and gene expression at gastrulation.

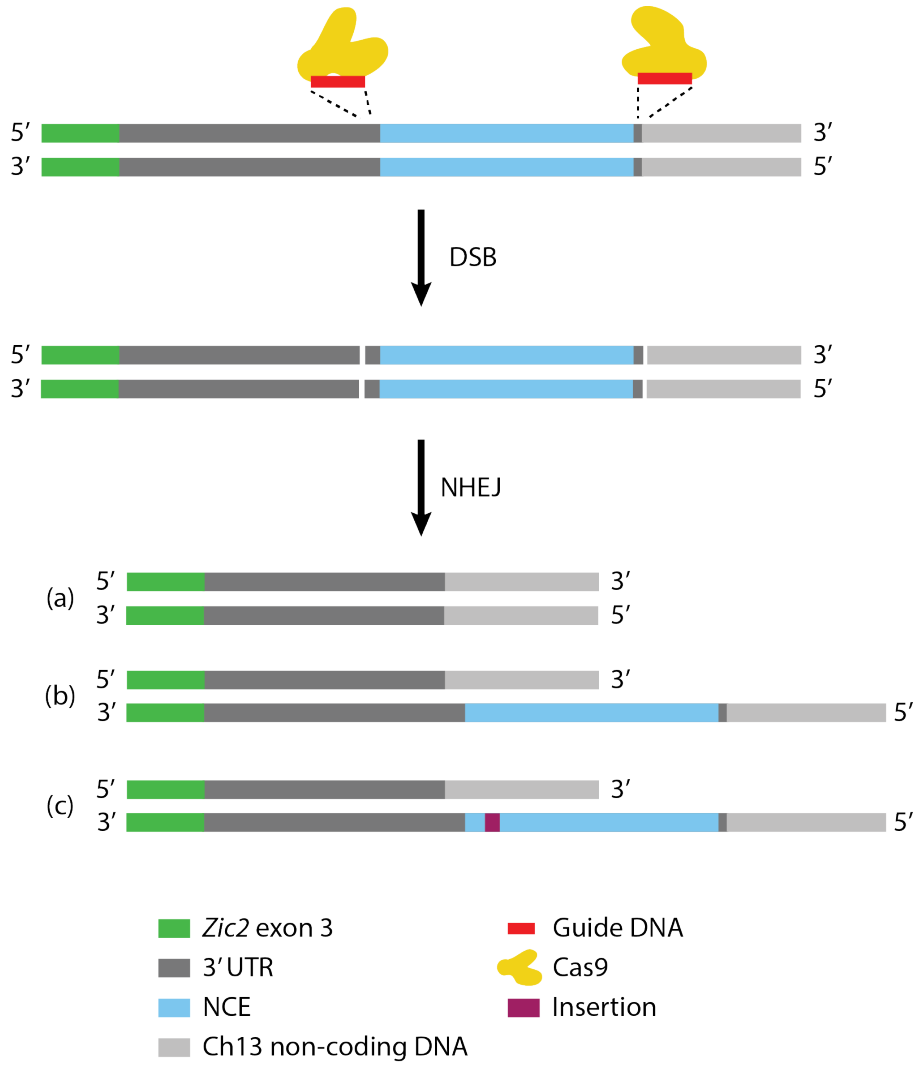


Figure 7.2: CRISPR *Zic2* NCE deletion via non-homologous end joining (NHEJ). Cas9 protein is targeted by guide DNA to induce double stranded breaks (DSB); in this example on either side of the *Zic2* NCE. Upon breaking, the *Zic2* NCE is released and the remaining DNA strands are ligated together via NHEJ. This can result in a number of different mouse genotypes, such as but not limited to **(a)** mutants homozygous for the NCE deletion, **(b)** mutants heterozygous for the NCE deletion, or **(c)** mutants heterozygous for the NCE deletion with additional insertions or point mutations.

7.2 Results

7.2.1 Generation of *PiggyBac* transgene constructs

PiggyBac transposon vectors (pBB232, pBB256, pBB259 and pBB262) were a gift from Bryan Bjork (Harvard Medical School, Boston, MA; Bjork et al. 2010). The properties of each construct are outlined in Table 7.1 (Appendix Figure A5.2). As pBB262 contained four copies of the cHS4 insulator sequence and thus the best protection from position effects during transgenesis, this vector was chosen for the initial *ZIC2* 3'UTR transgenesis experiments. Gateway® Entry vectors (pCR8-GW-TOPO-NCE) containing 608 bp of either the wildtype or mutated forms of the human *ZIC2* 3'UTR (Appendix Figure A5.3) were provided by the Muenke laboratory (NHGRI, NIH). The 608 bp 3'UTR fragment includes the entire 540 bp *ZIC2* NCE region and will be referred to as the NCE in all constructs.

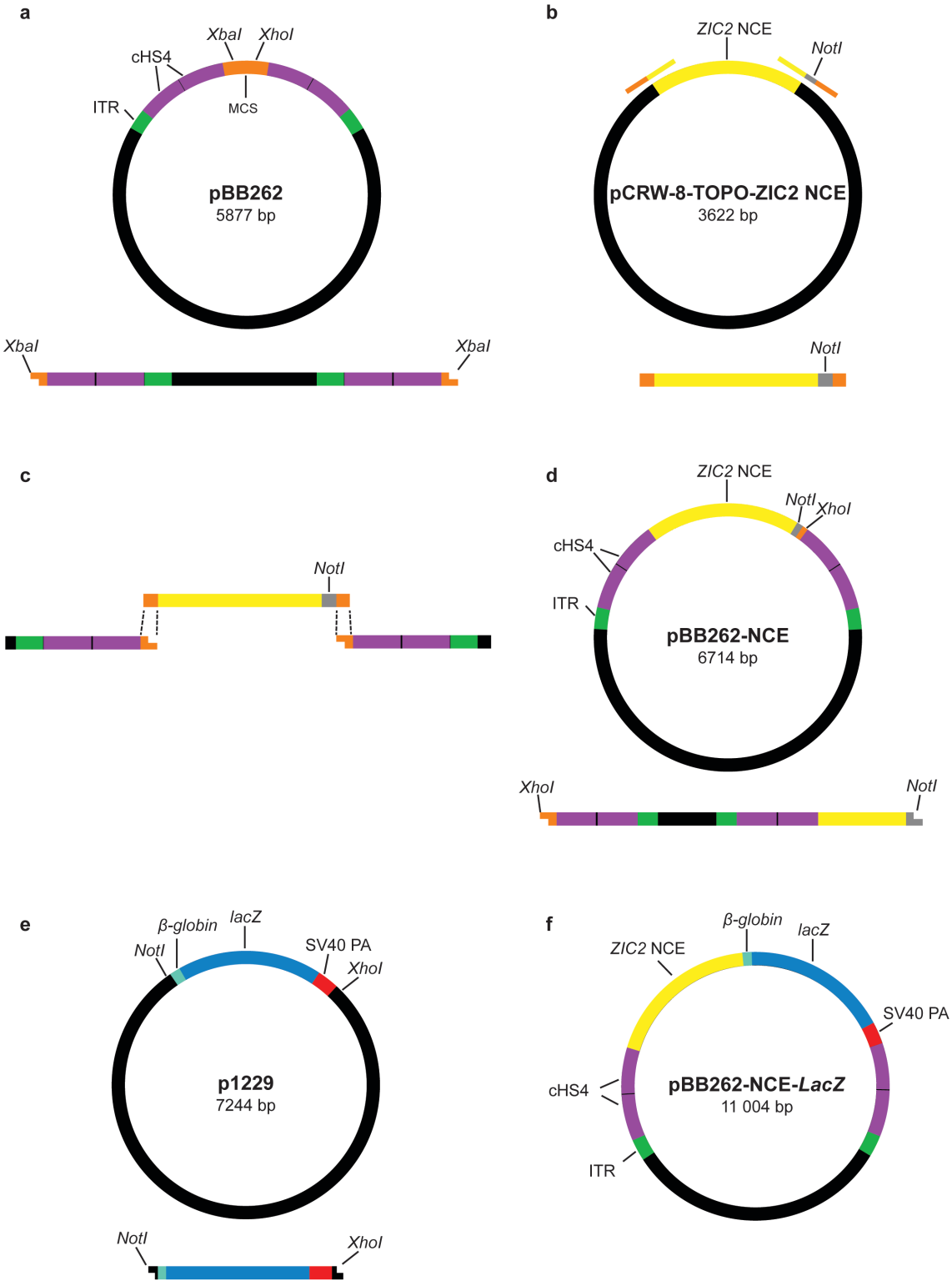
The 608 bp wildtype *ZIC2* NCE, and each of the six mutated versions of the *ZIC2* NCE, were PCR amplified from the pCR8-GW-TOPO-NCE constructs with oligonucleotides containing homology to the pBB262 MCS at the 5' ends and In-Fusion cloned into pBB262 that had been linearized with *Xba*I in the MCS (Figure 7.3, Chapter 2, Appendix Table A1.2, Appendix Figure A5.3). Following isolation and large-scale prepping, pBB262-*ZIC2* NCE constructs were digested with *NotI* and *XhoI*. In parallel, p1229, a Bluescript II KS vector containing a 4354 bp *lacZ* reporter/SV40 PolyA cassette was digested with *Not*I and *Xho*I to release the *lacZ*/SV40 PolyA. This was subsequently cloned into the digested pBB262-NCE constructs, resulting in wildtype and mutant pBB262-NCE-*lacZ* plasmids. A negative control was created by transferring the *lacZ* cDNA from p1229 to the MCS of pBB262, following digestion of both constructs with *Xba*I and *Xho*I, to generate pBB262-*lacZ*. A positive control was created via digestion of pmNeT-*lacZ*, a transgenic construct containing the mouse notochord enhancer for *Foxa2*. This construct had previously been used to create a conventional transgenic line (mNet) that exhibited *lacZ* reporter expression in the mouse node and notochord at 8.0-9.0 dpc (Ruth Arkell, unpublished; Figure 7.4). The original mNet mouse line at the ANU (Canberra, Australia) was used to confirm the X-Gal staining protocol was optimised for *lacZ* detection in gastrulating and post-gastrulation embryos (Figure 7.4), and to provide a comparison for future pBB262-mNet-*lacZ* embryos.

Prior to their use in mouse pronuclear injections, all *PiggyBac* constructs were tested for their ability to express the *lacZ* reporter via transfection into HEK293T cells. As can be seen in Figure 7.5, *lacZ* expression was detected 24 hours post transfection for all constructs containing the reporter gene, confirming construct integrity.

Table 7.1: Properties of the four *PiggyBac* transposon vectors. Vectors were gifted from Bryan Bjork (Harvard Medical School, Boston, MA; Bjork et al. 2010). Plasmid maps can be found in Appendix Figure A5.2. MCS: multiple cloning site; cHS4: chicken insulator fragment; ITR: inverted terminal repeats.

Plasmid	Purpose	ITR	cHS4 insulator sequences
pBB232 (pT3Ts-Pbase)	Contains the Pbase (*PiggyBac*) transposase open reading frame. *PiggyBac* is transcribed to mRNA and injected alongside pBB constructs containing the desired transgene cargo.	No	None
pBB256	Delivery of desired transgene cargo.	Yes	None
pBB259	Delivery of desired transgene cargo.	Yes	Two in total, one copy on either side of the MCS.
pBB262	Delivery of desired transgene cargo.	Yes	Four in total, two copies on either side of the MCS.

Figure 7.3: Generation of *ZIC2* NCE-*lacZ* *PiggyBac* constructs. (a) The MCS (orange) of the *PiggyBac* vector pBB262 was digested with *XbaI* to linearize the vector. pBB262 contains four cHS4 insulator fragments (purple) and two ITRs (green). **(b)** The *ZIC2* NCE fragment (yellow) was PCR amplified from pCR8-GW-TOPO-NCE constructs. The primers used to amplify the *ZIC2* NCE contained an artificial *NotI* site and regions homologous to the pBB262 MCS. **(c)** The ZIC-NCE amplicon and linearized pBB262 vector were joined via an In-Fusion (Clontech) reaction, **(d)** resulting in pBB262-NCE. pBB262-NCE was digested with *NotI* and *XhoI* to linearize the vector. **(e)** Simultaneously, p1229 was digested with *NotI* and *XhoI*, resulting in release of a *β-globin* minimal promoter-*lacZ*/SV40 PolyA cassette. **(f)** The *β-globin-lacZ*/SV40 PolyA cassette was subsequently ligated into pBB262-NCE to create pBB262-NCE-*lacZ* constructs. cHS4: chicken insulator fragments; ITR: inverted terminal repeats; MCS: multiple cloning site.


a
XbaI
XhoI
cHS4
ITR
MCS
pBB262
5877 bp
XbaI
XbaI
b
ZIC2 NCE
NotI
pCRW-8-TOPO-ZIC2 NCE
3622 bp
NotI
c
NotI
d
ZIC2 NCE
NotI
XhoI
cHS4
ITR
pBB262-NCE
6714 bp
XhoI
NotI
e
β-globin
lacZ
NotI
SV40 PA
XhoI
p1229
7244 bp
NotI
XhoI
f
β-globin
lacZ
ZIC2 NCE
SV40 PA
cHS4
pBB262-NCE-LacZ
11 004 bp
ITR

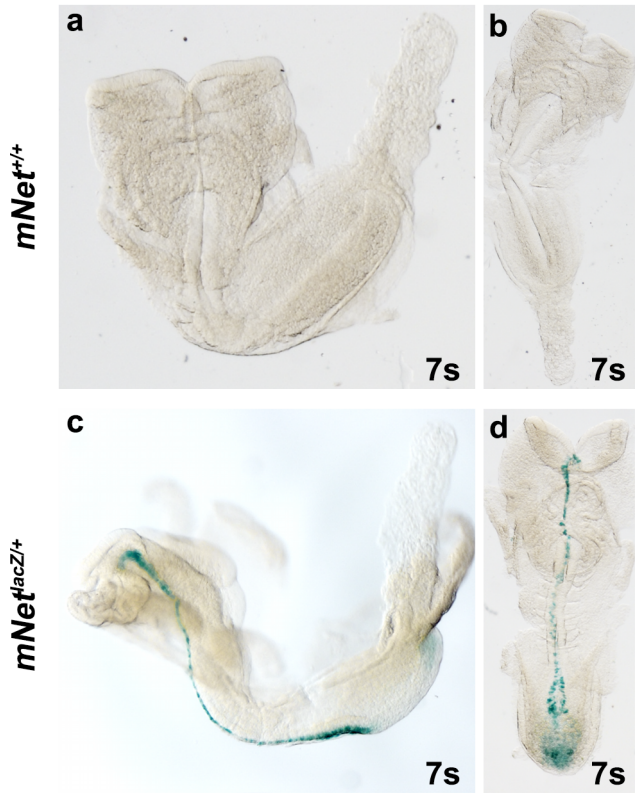


Figure 7.4: *lacZ* expression is detected in the node and notochord of mNet embryos. 7 somite stage embryos of the genotypes shown following X-Gal staining. **(a, c)** Lateral view; anterior is to the left. **(b, d)** Ventral view; anterior is to the top. *lacZ* expression is detected in the node and notochord of *mNet*$^{lacZ/+}$ embryos. s: somites.

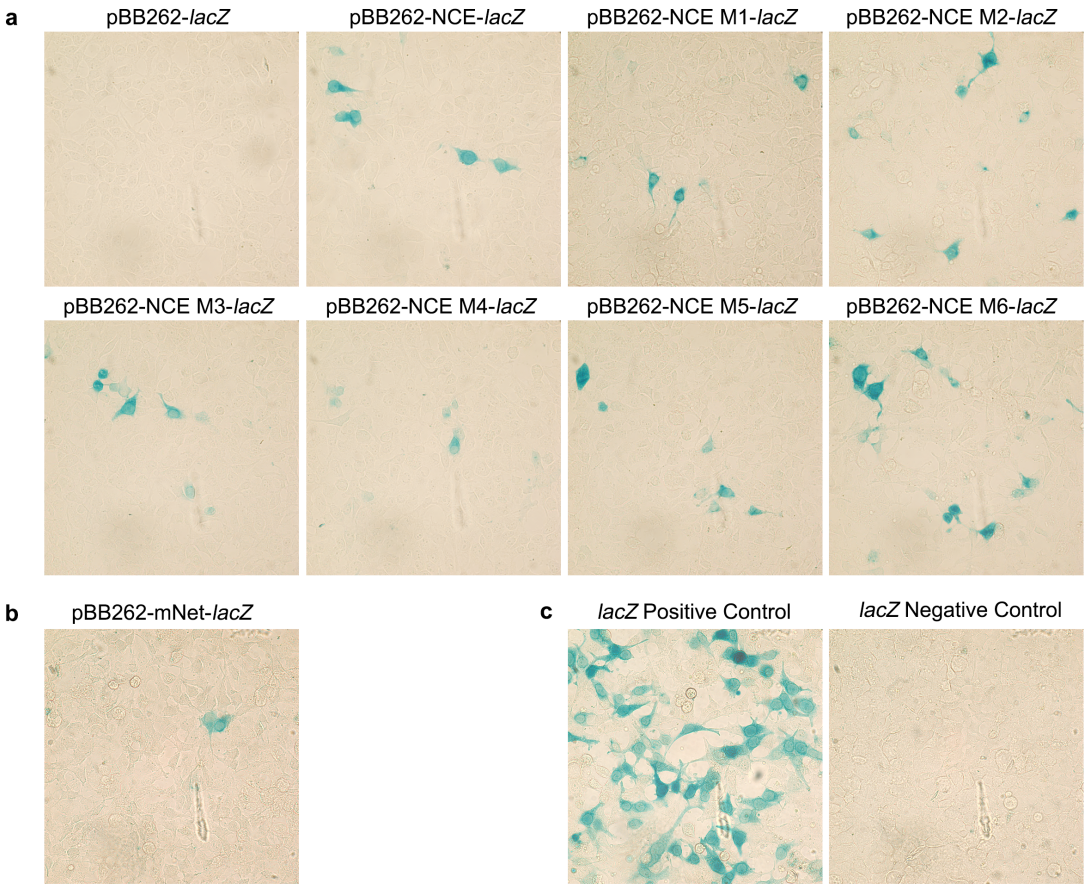


Figure 7.5: *PiggyBac* transposon constructs express *lacZ*. HEK293T cells were transfected with the plasmids indicated and stained with X-Gal to identify *lacZ* reporter expression. **(a)** Each *ZIC2* NCE *PiggyBac* construct, **(b)** the mNet *PiggyBac* construct, and the **(c)** pTracer-CMV/Bsd/*lacZ* produced *lacZ* expression in HEK293T cells. Expression was not detected in the negative control.

7.2.2 *Zic2* reference gene expression

If the NCE controls the expression of *Zic2* in vivo, *lacZ* reporter expression in the transgenic embryos is expected to recapitulate some or all domains of *Zic2* expression. *Zic2* expression at 6.5-7.75 dpc has been documented (Chapter 4, [Elms et al. 2004]), but *Zic2* expression at older ages has not. C3H/HeH wildtype embryos were dissected at 8.5 and 9.5 and WMISH for *Zic2* performed. As 12.5 dpc was the equivalent age in murine development to when GFP expression was detected in transgenic zebrafish embryos by Roessler et. al. (2012), *Zic2* expression was also examined at this stage. As shown in Figure 7.6, *Zic2* is expressed in the neural tube and headfolds at 8.5 dpc, and the neural tube, fore and mid-brain, and developing forelimbs at 9.5 dpc. At 12.5 dpc, *Zic2* appears to be expressed in the diencephalon, midbrain and hindbrain, along with the midline and somites/developing vertebrae. Expression can also be detected in the apical ectodermal ridge (AER) and zone of polarizing activity (ZPA) of the developing fore- and hindlegs. Sectioning of the embryos is required to confirm the exact distribution of transcripts, however.

7.2.3 *PiggyBac* is a highly efficient method for the creation of transgenic embryos

Prior to dissection of *PiggyBac* transgenic embryos, a positive control was required to batch test each round of X-Gal stain at the National Institute of Health (Maryland, USA) where the PiggyBac experiments were being conducted. As the original mNet line was at the ANU, a $Sox10^{lacZ}$ strain (Britsch *et al.*, 2001) already present at the NIH Transgenic Core Facility was chosen. $Sox10^{lacz/+}$ females were timed mated with wildtype ($Sox10^{+/+}$) males. Embryos were dissected at 14.5 dpc and, along with ear notches from $Sox10^{lacz/+}$ females, stained in X-Gal overnight. Blue staining in both the embryos and ear notches indicated that the batch of X-Gal was sufficient for use on *PiggyBac* transgenic embryos (Appendix Figure A5.4).

To create transient transgenic *ZIC2* NCE-*lacZ* embryos, *PiggyBac* transposon mRNA and pBB262 transgene constructs (pBB262-mNet-*lacZ*, pBB262-*lacZ* and pBB262-NCE-*lacZ*) linearized with *XmnI* were co-injected into the pronucleus by Gene Elliot at the NIH Transgenic Core Facility. A total of 19 transgenesis rounds were conducted, with the concentrations of transposase mRNA and reporter DNA for each injection outlined in Table 7.2. $Tn^{(pb\text{-pBB262-mNet-}lacZ)Ark}$ transgenic embryos were dissected at 8.5 dpc as this is the known time of mNet enhancer activity, whilst $Tn^{(pb\text{-pBB262-NCE-}lacZ)Ark}$ and $Tn^{(pb\text{-pBB262-}lacZ)Ark}$ embryos were dissected at 7.5 dpc, 9.5 dpc or 12.5 dpc. All recovered embryos were genotyped for the presence of *lacZ*. In total, 23 embryos were recovered following pBB262-mNet-*lacZ* pronuclear injection and 261 embryos were recovered following pBB262-NCE-*lacZ* injection. As shown in Table 7.2, the concentration of DNA and mRNA injected appeared to influence the rate of transgenesis for pBB262-NCE-*lacZ* (rounds 1-2 versus rounds 5-14) but not pBB262-mNet-*lacZ* (rounds 3-4 versus

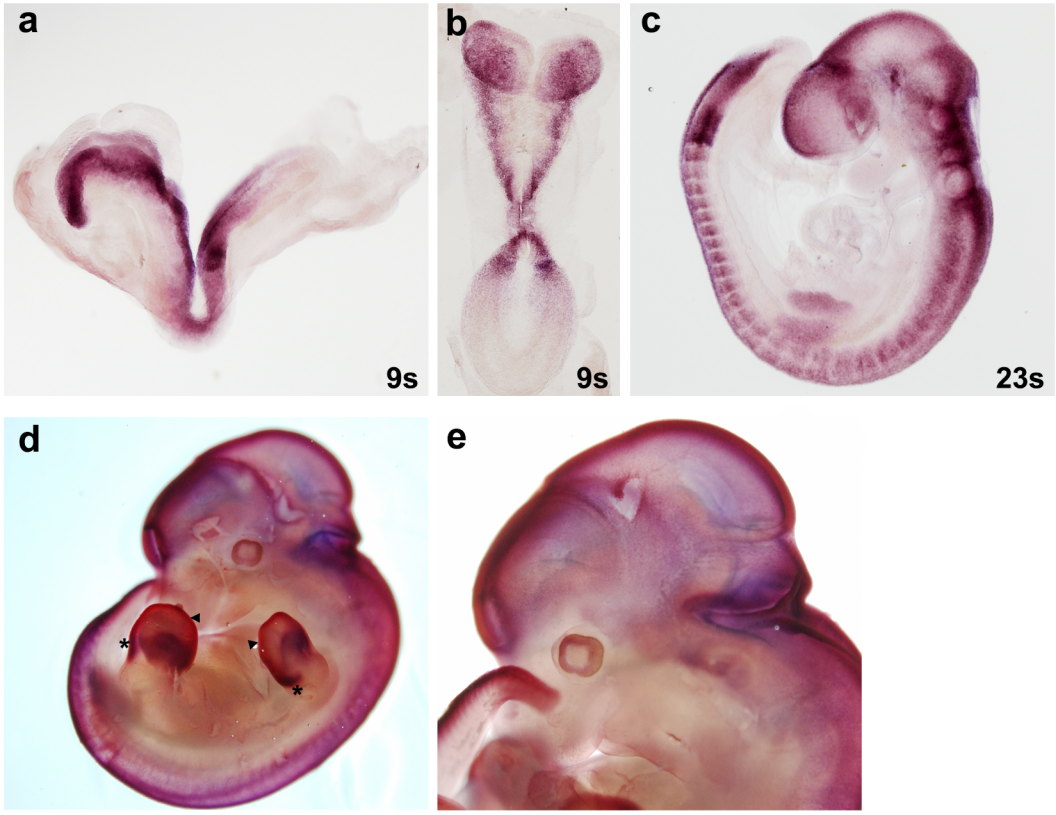


Figure 7.6: *Zic2* expression in post-gastrulation murine embryos. (a, c-e) Lateral (anterior to the left) and **(b)** ventral (anterior to the top) view of wildtype embryos after WMISH to *Zic2*. **(a-b)** *Zic2* appears to be expressed in the headfold neurectoderm at 8.5 dpc, **(c)** in the neural tube of the fore- and mid-brain, and developing forelimb at 9.5 dpc, **(d-e)** and in the diencephalon, midbrain and hindbrain, midline, developing vertebrae of the spine, apical ectodermal ridge (arrowheads) and zone of polarizing activity (*) of the developing fore- and hind-legs. S: somites.

Table 7.2: *PiggyBac* transgenic injections. The expected number of *lacZ*/+ embryos was calculated as 68% of the total number of viable embryos, as this was the value reported in the literature (Bjork *et al.*, 2010). Number of *lacZ*/+ embryos: the number of embryos which genotyped positive for the *lacZ* transgene, indicating transgenesis was successful. Transgenesis rate was calculated as the percent of embryos that genotypes positive for the *lacZ* transgene out of the total number of viable embryos. Resorption rate was calculated as the percentage of empty decidua out of total decidua. ND: not determined. PsP: pseudopregnant.

Injection round No.	No. of PsP females	Plasmid DNA (ng/μL)	*PiggyBac* mRNA (ng/μL)	No. of decidua	Total No. viable embryos	Expected No. *lacZ*/+ embryos	No. *lacZ*/+ embryos	No. *lacZ* expressing embryos	Embryo age	Transgenesis rate (%)	Resorption rate (%)
					pBB262-NCE-*lacZ*						
1	10	4	34	23	14	9.5	0	0	7-7.5	0.0	39.1
2	7	4	34	28	18	12.2	2	0	7-7.5	11.1	35.7
5	13	2	23	52	31	21.1	25	0	6-7.5	80.6	40.4
6	9	2	23	30	10	6.8	8	0	7-7.5	80.0	66.7
7	7	2	23	35	27	18.4	9	0	7-7.5	33.3	22.9
9	11	2	23	40	19	12.9	10	0	7-7.5	52.6	52.5
10	17	2	23	59	23	15.6	17	0	8.5-9.5	73.9	61.0
11	8	2	23	22	9	6.1	4	1	12.5	44.4	59.1
13	3	2	23	16	12	8.1	4	0	12.5	33.3	25.0
14	5	2	23	11	6	4.0	2	1	8.5	33.3	45.5
15	5	2	23	9	2	1.4	0	0	9.5	0.0	77.8
16	5	2	23	27	18	12.2	ND	0	7.5	ND	33.3
17	13	2	23	49	18	12.2	ND	0	9.5	ND	63.3
18	10	2	23	40	18	12.2	ND	0	8.5	ND	55.0
19	10	2	23	64	36	24.5	ND	3	12.5	ND	43.8
Total:	**133**			**505**	**261**	**177.4**	**81**	**5**	**Avg:**	**40.2**	**48.1**
					pBB262-mNet-*lacZ* (positive control)						
3	6	4	34	12	8	5.4	7	0	8-8.5	87.5	33.3
4	6	4	34	21	11	7.5	10	0	8-8.5	90.9	47.6
8	6	2	23	11	4	2.7	4	0	8-8.5	100.0	63.6
Total:	**18**			**44**	**23**	**15.6**	**21**	**0**	**Avg:**	**92.8**	**48.2**

round 8), suggesting the transgenesis rate was construct dependant. Overall, the average transgenesis rates for both constructs (92.8% for pBB262-mNet-*lacZ* and 40.2% for pBB262-NCE-*lacZ;* Table 7.2) suggests the *PiggyBac* transposase is highly effective at introducing transgenes into the murine genome, compared to standard transgenesis where ~2-5% of microinjected pronuclei are recovered as transgenic founders (Nakanishi *et al.*, 2002; Wall, 2001).

7.2.4 *lacZ* expression could not be detected in transgenic embryos

It was expected that the number of *lacZ* expressing embryos would correlate with the number of embryos that genotyped positive for the *lacZ* reporter. Despite detecting a medium-high level of transgenesis in both Tn$^{(pb\text{-pBB262-mNet-}lacZ)\text{Ark}}$ embryos and Tn$^{(pb\text{-pBB262-NCE-}lacZ)\text{Ark}}$ embryos, only 6% of embryos that genotyped positive for the pBB262-NCE-*lacZ* transgene exhibited *lacZ* expression when stained with X-Gal (Figure 7.7, Appendix Figure A5.5). This small expression rate is consistent with the NCE acting as a transcriptional repressor (Chapter 5). Moreover, the *lacZ* expression patterns detected in the positive embryos were variable, suggesting that it is not a true readout of NCE activity. However, 0% of positive control embryos that genotyped positive for pBB262-mNet-*lacZ* exhibited *lacZ* expression, suggesting that the cHS4 insulators within the transgene were overriding the promoter and supressing *lacZ* expression. This lack of positive control expression prevented meaningful interpretation of the experimental transgene. As such, the six HPE-associated SNV constructs (M1-M6) were not tested and no pBB262-*lacZ* (negative control) injections were performed.

7.2.5 Optimization of *Zic2* 3'UTR CRISPR-Cas9 mutagenesis

Design of the CRISPR guide strands, as well as pronuclei injection and implantation, and genotyping of pups, was carried out in conjunction with the Transgenesis Facility (JCSMR, ANU, Canberra). Eight guides targeting the *Zic2* 3'UTR were initially designed and selected using the algorithms available via the Zhang CRISPR design site (Zhang Lab and MIT, 2015) (Table 7.3, Figure 7.8). Of the eight selected guides (Table 7.3), two that were situated either side of the *Zic2* NCE (designated G2 and G3) were initially chosen for the mutagenesis experiments. G2 and G3 DNA sequences were cloned into the guide plasmid constructs and the isolated plasmid incorporated into pronuclei of fertilized zygotes (either by injection or electroporation) (Appendix Figure A5.6). These experiments formed part of the Transgenesis Facility's optimisation of CRISPR-Cas9 protocols, resulting in inter-experimental variation in the method of zygote incorporation, the form in which reagents were used, and the concentration of reagents (Table 7.4 and Table 7.5). The stage at which the transferred embryos developed to before analysis also varied. A total of 25 CRISPR-Cas9 rounds sessions, in which ~20 2-cell embryos were implanted per dam, yielded only three embryos with altered sequence at the

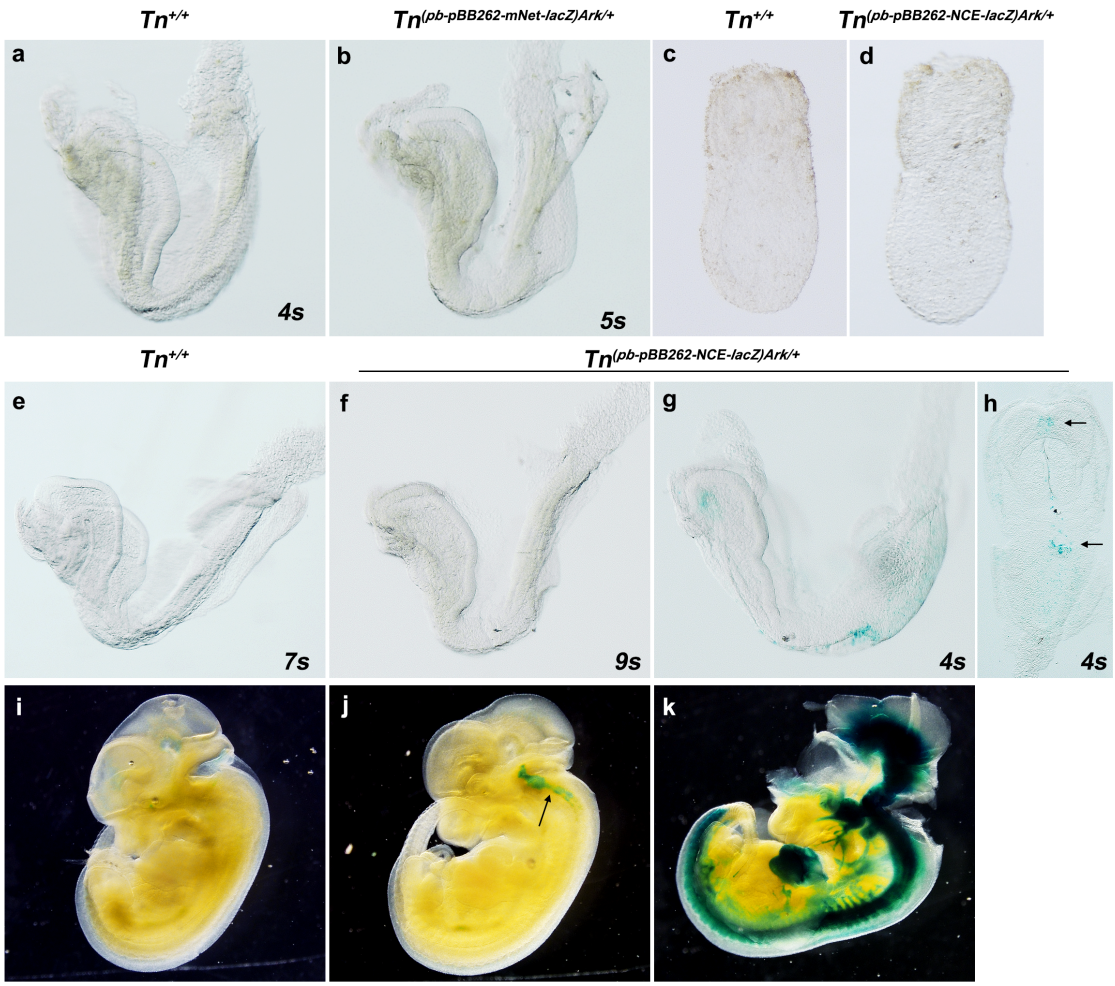


Figure 7.7: *PiggyBac* transgenic embryos. Lateral view of embryos of the genotypes shown following X-Gal staining. **(a)** A $Tn^{+/+}$ and **(b)** $Tn^{(pb\text{-}pBB262\text{-}mNet\text{-}lacZ)Ark/+}$ embryo (positive control) for the stages shown. Anterior is to the left. No *lacZ* expression was detected in any $Tn^{(pb\text{-}pBB262\text{-}mNet\text{-}lacZ)Ark/+}$ embryo. Ventral view of an **(c)** early-streak embryo and a **(d)** mid-streak $Tn^{(pb\text{-}pBB262\text{-}NCE\text{-}lacZ)Ark/+}$ embryo of the genotypes shown following X-gal staining. Anterior is to the left. No *lacZ* expression was detected. Lateral view of **(e)** $Tn^{+/+}$ and **(f-g)** $Tn^{(pb\text{-}pBB262\text{-}NCE\text{-}lacZ)Ark/+}$ 8-8.5 dpc embryos following X-Gal staining. Anterior is to the left. **(h)** Ventral view. One $Tn^{(pb\text{-}pBB262\text{-}NCE\text{-}lacZ)Ark/+}$ embryo **(g-h)** exhibited *lacZ* expression in the node and midbrain (black arrows). Lateral view of **(i)** $Tn^{+/+}$ and **(j-k)** $Tn^{(pb\text{-}pBB262\text{-}NCE\text{-}lacZ)Ark/+}$ embryos at 12.5 dpc following X-gal staining. Anterior is to the left. **(j)** X-gal stain trapping was detected in some $Tn^{(pb\text{-}pBB262\text{-}NCE\text{-}lacZ)Ark/+}$ embryos (black arrow), but this was not true *lacZ* expression. Four $Tn^{(pb\text{-}pBB262\text{-}NCE\text{-}lacZ)Ark/+}$ embryos **(k, Appendix Figure A5.5)** exhibited *lacZ* expression, however the expression pattern was inconsistent between embryos. S: somites.

Table 7.3: Selected *Zic2* 3'UTR CRISPR-mutagenesis guides. Guides were assessed via the following criteria: a score of 1-20 is considered a low-quality guide, 20-49 a mid-quality guide, and ≥50 a high-quality guide (Zhang Lab and MIT, 2015). 'Off target' denotes the total number of matching sites for the guide within the genome, with brackets indicating the number of off target sites occurring within coding regions only. 'Mismatches' represent the number of non-paired bases a the predicted off target sites. Preferred guides have a score of ≥50, ≥2 mismatches and a minimal number of off-target sites. PAM: Protospacer Adjacent Motif.

Type	Guide Name	Guide Sequence 5'-3'	Guide *Zic2* 3'UTR Location (bp)	PAM sequence	Score	Off Target	Mismatches
DNA	G2	AGCAGAGTTATTGAAGAGAATGG	1108-1131	TGG	34	500 (25)	1-4
DNA	G3	CGGTGATTTTAATGGCTTATTGG	356-379	TGG	72	144 (3)	3-4
RNA	G5	CGCCCTCCCAAAACCCATCG	43-63	AGG	88	77 (14))	3-4
RNA	G6	AAACCCATCGAGGGCACCTT	53-73	AGG	81	87 (15)	3-4
RNA	G7	AAAAAGATTTTACCAGCAGA	119-139	AGG	44	500 (27)	2-4
RNA	G8	TTTTGACAAACTGTACATAG	317-337	CGG	65	326 (17)	3-4
RNA	G9	ATGGCTTATTGGCTTATTGG	367-387	TGG	71	175 (12))	2-4
RNA	G10	AACTTCATGCGACTATAGAC	657-677	TGG	83	73 (6)	2-4

Figure 7.8: Location of CRISPR-mutagenesis guides within the *Zic2* 3'UTR. Guide strands (orange) were designed to target multiple locations within the *Zic2* 3'UTR. Black: six human HPE-associated SNVs, green: putative ARE region, pink: predicted poly(A) sites.

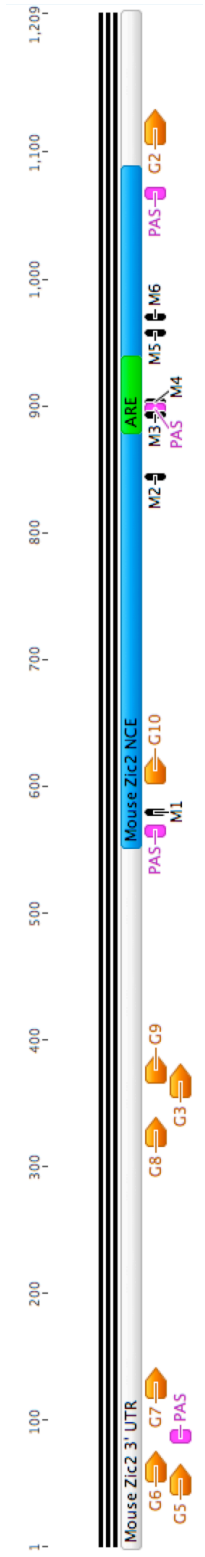


Table 7.4: Rounds of *Zic2* 3'UTR CRISPR mutagenesis performed via microinjection. Viability from 2 cell (%): the percentage of viable embryos when dissected out of the total number of implanted 2 cells. Resorption rate (%): the percentage of non-viable embryos when compared to total number of decidua. The preparation of the Cas9 protein, guide strands, zygote harvesting and microinjection were all performed by the JCSMR Transgenesis Facility (ANU, Canberra). ND: not determined; plDNA: plasmid DNA; WT: wildtype.

		CRISPR Settings				Dissection					
	Round	Guide type	Guides	Total Guide concentration (ng/μL)	Cas9 protein (ng/μL)	No. of 2-cell embryos implanted	Dissection age (dpc)	No. decidua	No. viable progeny	Resorption rate (%)	Viability from 2-cell embryos (%)	Genotyping results
	1	plDNA	G2/G3	60	60	ND	6.5	9	5	44.44	ND	All WT
	2	plDNA	G2/G3	60	60	ND	7.5	1	0	100.00	ND	-
	3	plDNA	G2/G3	60	60	ND	9.5	5	3	40.00	ND	All WT
	4	plDNA	G2/G3	60	60	ND	9.5	9	5	44.44	ND	All WT
	5	plDNA	G2/G3	30	50	ND	9.5	4	1	75.00	ND	All WT
	6	plDNA	G2/G3	30	50	ND	9.5	11	7	36.37	ND	All WT
	7	plDNA	G2/G3	2.5	50	ND	8.5	11	6	45.45	ND	Two putative mutants
	8	plDNA	G2/G3	2.5	50	21	7.5	3	0	100.00	0.00	-
	9	plDNA	G2/G3	2.5	50	21	7.5	9	6	33.33	28.57	All WT
	10	plDNA	G2/G3	2.5	50	19	7.5	15	3	80.00	15.78	All WT
	11	plDNA	G2/G3	2.5	50	20	7.5	10	5	50.00	25.00	All WT
	25	plDNA	G2/G3	2.5	50	21	Littered Down		1	95.23	4.77	All WT
	26	mRNA	G6/G7/G8	2.5	50	53	Littered Down		8	84.91	15.09	Pups eaten before genotyped
	27	mRNA	G6/G7/G8	2.5	50	42	Littered Down		3	92.86	7.14	Three mice with putative 3'UTR deletions

Table 7.5: Rounds of *Zic2* 3'UTR CRISPR mutagenesis performed via electroporation. Viability from 2 cell (%): the percentage of viable embryos when dissected out of the total number of implanted 2-cells. Resorption rate (%): the percentage of non-viable embryos when compared to total number of decidua. The preparation of the Cas9 protein, guide strands, zygote harvesting and electroporation were all performed by the JCSMR Transgenesis Facility (ANU, Canberra). V/ms/pulse: volts, milliseconds, number of pulses; ND: not determined; plDNA: plasmid DNA; WT: wildtype.

		CRISPR Settings					Dissection					
	Round	Guide type	Guide concentration (ng/μL)	Cas9 protein (ng/μL)	Electroporation settings (V/ms/pulse)	No. of 2-cell embryos implanted	Dissection age (dpc)	No. decidua	No. viable progeny	Resorption rate (%)	Viability from 2-cell embryos (%)	Genotyping results
	12	plDNA	2.5	50	700/10/1	19	7.5	13	6	53.84	31.57	All WT
	13	plDNA	2.5	50	700/20/1	20	7.5	11	6	45.45	30.00	All WT
	14	plDNA	2.5	50	800/10/1	37	7.5	9	5	44.44	13.52	One putative mutant
	15	plDNA	2.5	50	800/10/1	20	7.5	13	9	30.70	45.00	All WT
	16	plDNA	2.5	50	800/10/1	20	7.5	3	1	66.67	5.00	All WT
	17	plDNA	2.5	50	800/10/1	23	7.5	13	8	38.46	34.78	All WT
	18	plDNA	2.5	50	800/10/1	23	7.5	17	11	35.29	47.82	All WT
	19	plDNA	50.0	100	800/10/1	24	Littered Down		3	87.50	12.50	All WT
	20	plDNA	50.0	100	800/10/1	24	Littered Down		0	100.00	0.00	-
	21	plDNA	25.0	150	800/10/1	23	Littered Down		0	100.00	0.00	-
	22	plDNA	25.0	150	800/10/1	23	Littered Down		0	100.00	0.00	-
	23	plDNA	2.5	50	800/10/1	21	Littered Down		3	85.71	14.29	All WT
	24	plDNA	2.5	50	800/10/1	21	Littered Down		0	100.00	0.00	-

NCE. To exclude the possibility that modification of the NCE leads to in utero lethality, genotyping of microinjected 2-cell embryos and sequencing of the genotyping amplicons was performed. The *Zic2* 3'UTR was wildtype in all embryos examined (n=10, data not shown), excluding lethality as the cause of the low mutagenesis frequency and instead indicating a low frequency of DNA breaks induced by G2 and G3.

To increase the likelihood of the selected guides generating in vivo cuts, the remaining guides were synthesised as RNA, enabling their cutting ability to be tested in vitro via a cleavage assay. As shown in Figure 7.9 and Table 7.7, G6, G7, G8 and G10 appeared to induce Cas9 to the PCR target. When G6, G7 and G8 were co-injected into pronuclei with Cas9 protein, only a small number of pups were recovered (Figure 7.10). All three pups, however, carried a variant sequence at the *Zic2* 3'UTR. Additional guides specifically targeting the NCE have since been designed and are being used in the current rounds of *Zic2* 3'UTR CRISPR-Cas9 mutagenesis.

7.2.6 Analysis of *Zic2* 3'UTR CRISPR-Cas9 embryos and mice

In the 25 rounds of CRISPR-Cas9 mutagenesis performed with G2 and G3, three embryos were identified as possible mutants via genotyping. Those embryos at the mid-streak stage were analysed via WMISH to *Zic2* and compared to wildtype littermates, whilst those at early somite stages were analysed with WMISH to *SHH*. Each modified embryo showed expression differences relative to control littermates (Figure 7.11). Further attempts to document the precise *Zic2* 3'UTR modification in each embryo via Sanger sequencing were not successful.

From the two CRISPR-mutagenesis rounds using G6, G7, and G8, three pups were found to have putative mutations within the *Zic2* 3'UTR via Sanger sequencing of the genotyping amplicon. The first, designated F0#8, exhibited a 1 bp deletion and a 2 bp deletion in the *Zic2* 3'UTR, upstream of the NCE (Figure 7.10). The second, F0#9, exhibited a 267 bp deletion and an overlapping 2 bp insertion in the *Zic2* 3'UTR, upstream of the NCE, whilst the third, F0#10, exhibited a deletion of unknown size in the same region. No visible phenotype characteristic of other *Zic2* heterozygous mutant strains (curly tail or belly spot) was seen in any of the mice.

Due to the presence of large deletions, F0#9 and F0#10 were retained for further analysis. Their progeny will be genotyped and Sanger sequenced to confirm the presence and exact location of the *Zic2* 3'UTR deletions, and phenotyped for *Zic2*-associated defects such as HPE, LR randomization and spina bifida. As the deletions occur upstream of the NCE region, it is expected that homozygote pups from these lines will not exhibit HPE. If this is correct, these lines can act as a negative control for any *Zic2* NCE deletions that arise from the ongoing CRISPR-Cas9 mutagenesis experiments.

Figure 7.9: Four *Zic2* 3'UTR guides sufficiently induced Cas9 cutting in an *in vitro* cleavage assay. A cleavage assay was performed by Nay Chi (JCSMR Transgenesis Facility) on G5-G10. A 730 bp *Zic2* 3'UTR PCR amplicon was incubated Cas9 protein and individual guides made up of crRNA annealed to tracrRNA. **(a)** Successful cleavage of the amplicon was determined by the production of two distinct products at the expected size. **(b)** When analysed on a 1.5% agarose gel, G6, G7, G8 and G10 were found to successful cleave the *Zic2* 3'UTR amplicon (*). L: HyperLadder 50 bp (Bioline).

A

Guide	UTR amplicon	Band 1 expected size	Band 2 expected size
G5	730 bp	99 bp	631 bp
G6	730 bp	109 bp	621 bp
G7	730 bp	175 bp	555 bp
G8	730 bp	373 bp	357 bp
G9	730 bp	423 bp	307 bp
G10	730 bp	659 bp	71 bp

B

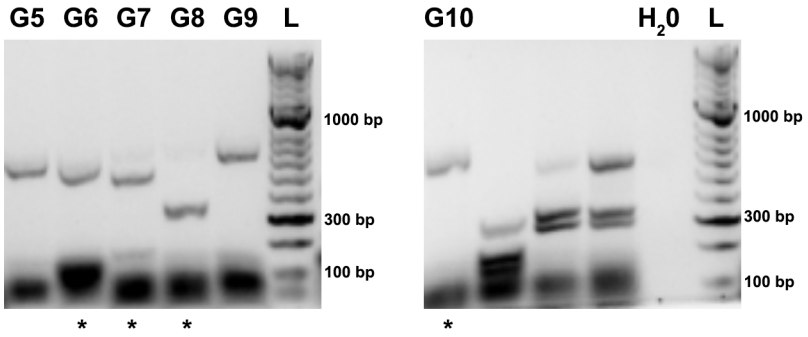


Figure 7.10: Sequence analysis of founder animals with putative deletions in the *Zic2* 3'UTR. **(a)** Sanger sequencing traces for three founder (F0). A double indicates a deletion. **(b)** Location of putative mutations in the *Zic2* 3'UTR for the corresponding founder mice.

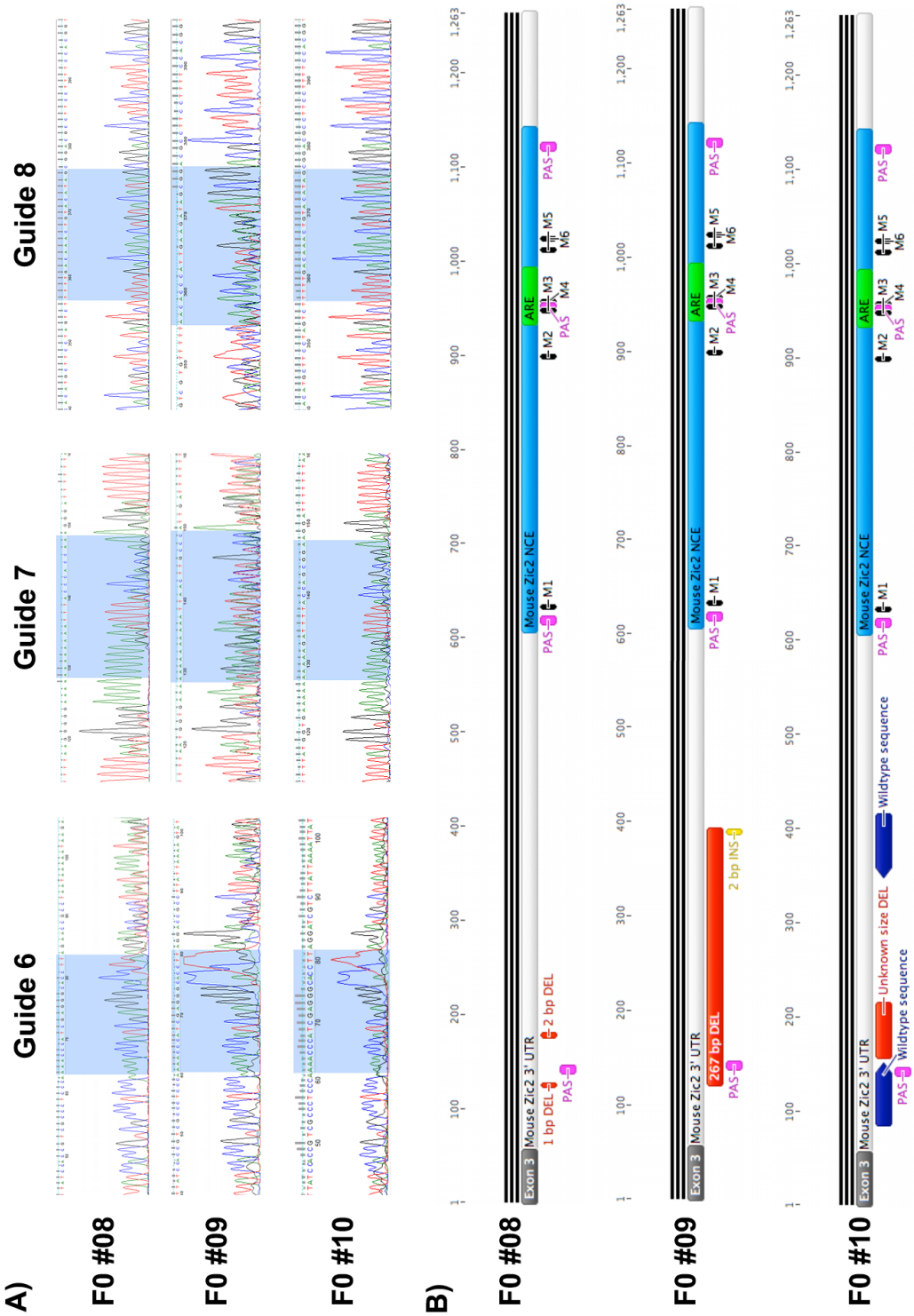
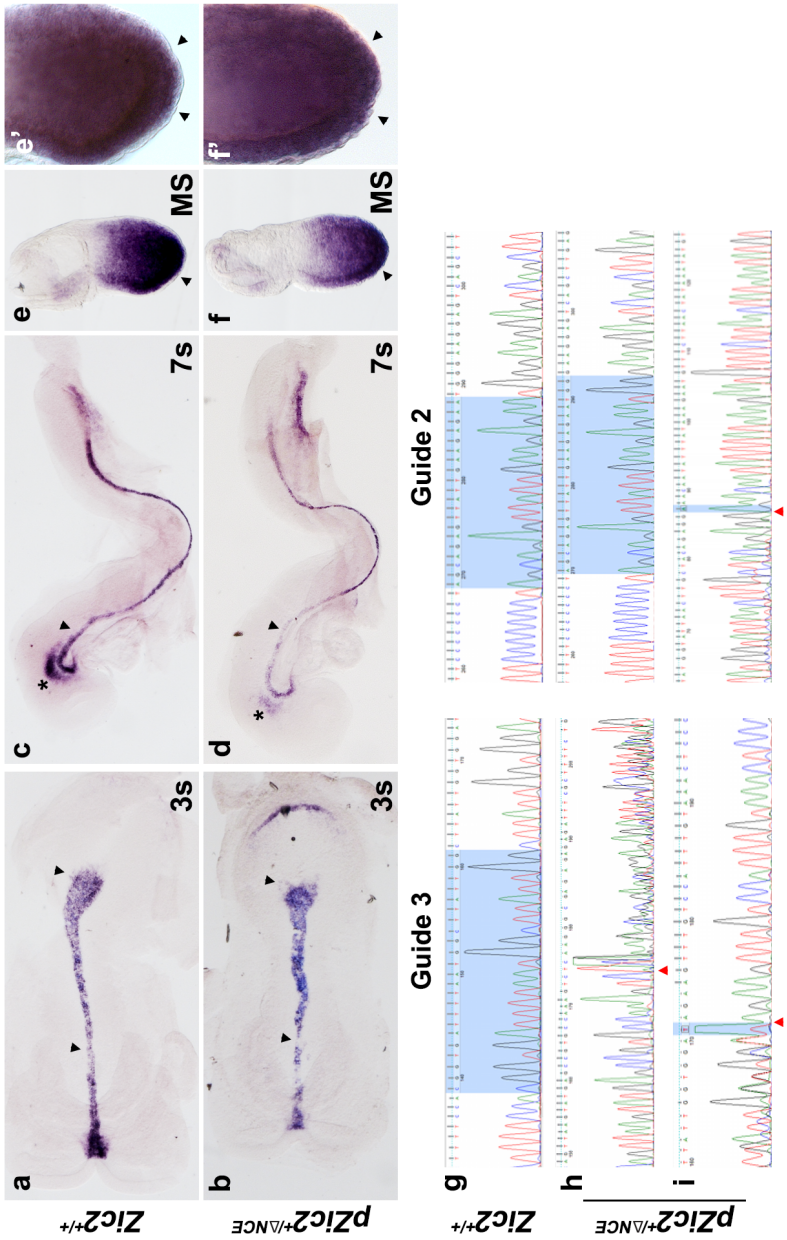


Figure 7.11: Analysis of unconfirmed *Zic2* NCE mutant embryos. *pZic2*$^{+/\Delta NCE}$ represents putative *Zic2* NCE mutants. Anterior is to the left. **(a-d)** WMISH to *Shh.* **(a-b)** Ventral view of 3s embryos of the genotypes shown. The *pZic2*$^{+/\Delta NCE}$ embryo exhibited a discontinued notochord and expanded node staining (black arrowheads). **(c-d)** Lateral view of 7s embryos of the genotypes shown. The *pZic2*$^{+/\Delta NCE}$ embryo exhibited a discontinued notochord (black arrowhead) and reduced *Shh* expression (*). **(e-f')** Lateral view of mid-streak stage embryos of the genotypes shown with WMISH to *Zic2*. The *pZic2*$^{+/\Delta NCE}$ embryo appeared to have either expanded *Zic2* in the node and surrounding endoderm, or a reduced endoderm layer (black arrowhead). **(g-i)** Examples of Sanger sequencing traces for embryos of the genotypes shown. Mutations/inconclusive reads are annotated by a red arrowhead. S: somites, MS: mid-streak.



7.3 Discussion

The aim of this chapter was to investigate the role of the *Zic2* NCE, using both gain-of-function (*PiggyBac* transposon transgenics) and loss-of-function (CRISPR-Cas9 mutagenesis) approaches, to determine if the NCE regulated *Zic2* expression during gastrulation and if it played a role in HPE pathogenesis. Whilst *PiggyBac* was shown to be a highly effective method for creation of transgenic embryos, this did not translate to reproducible NCE driven *lacZ* expression at gastrulation. Similarly, whilst CRISPR-Cas9 was used to successfully delete a portion of the *Zic2* 3'UTR, this region was not within the NCE.

7.3.1 Technical difficulties prevented the assessment of *ZIC2* UTR function via either gain- or loss-of-function methods

The *PiggyBac* transgenesis system led to a high frequency of transgenesis, but a low frequency of reporter expression. For $Tn^{(pb\text{-}pBB262\text{-}NCE\text{-}lacZ)Ark}$ embryos, this may be due to the *Zic2* NCE acting as a repressor at the stages analysed, as suggested by heterologous reporter assays in HEK293T cells (Chapter 5). The mNet-*lacZ* positive control also failed to produce reporter expression. One factor that may have prevented expression of the *lacZ* reporter in all constructs is over-insulation from the cHS4 fragments within the transgene. The cHS4 insulator fragments in the pBB262 plasmids were included to minimize position effects upon the inserted transgenes from adjacent regulatory elements and to prevent post-integration silencing, but it is possible that their enhancer-blocking properties blocked the activity of the *β-globin* promoter and/or NCE and mNet element activity, resulting in no or inconsistent *lacZ* expression. Whilst interference between cHS4 insulators and promoter-enhancer activity has previously been reported (Groth *et al.*, 2013), these effects were on activation of endogenous promoters in the host genome. Though the *β-globin* minimal promoter used in all *PiggyBac* experiments reported in this thesis was the same promoter used in the initial creation of the mNet line (Ruth Arkell, unpublished) indicating it can be utilised during mouse gastrulation, and the initial optimisation experiments by Aker *et. al.* (2007) for the cHS4 insulators used in this Chapter were conducted with a *β-globin* minimal promoter, it remains possible that the insulator-promoter combination used in these experiments does not drive robust expression at the specific stages analysed. To investigate this possibility, a series of transgenics made plasmids with varying numbers of, or no, cHS4 insulator fragments (Table 7.1) should be tested. Alternatively, different promoters known to be active at gastrulation could be incorporated into the pBB262-NCE-*lacZ* construct.

Another possibility is that the transgenesis protocol is genotoxic. This could occur due to chance integration at, and interruption of, an essential gene within the murine genome, or due to unstable transgene integration and continued genome mobility that can lead to genome rearrangements such as those previously been reported with *PiggyBac* transgenics (Huang *et*

al., 2010; Yusa *et al.*, 2011). One method to control for this possibility is to use a single plasmid construct that combines both the transposase and transpositioning transgene. By sharing a single poly(A) sequence, the transposase self-inactivates after the initial transposition event, preventing further events from occurring (Chakraborty *et al.*, 2014). Genotoxicity is expected to result in a decrease transgenic efficiency only, however, and is unlikely to account for the complete lack of *lacZ* expression in the positive control.

In addition to issues with the *PiggyBac* constructs, it is possible that variation in genetic background could account for the mNet-*lacZ* transgene successfully expressing *lacZ* in previous experiments but not in the transgenics reported here, as multiple studies report wide ranges of phenotypic variation dependant on the mouse strain used (Doetschman, 2009; Schachter and Krauss, 2008). The expected variation is minimal, however, since the original mNet line and the *PiggyBac* transgenics reported in this Chapter were produced on the same inbred strain (FVB), but different sub-strains. It is mostly likely a combination of the events discussed above, however, that resulted in a failure of the *PiggyBac* transgenics. As such, optimization and redesign will need to take place before further transgenics are attempted. Whilst conventional transgenics with DNA cargo 1-2 kb in size are still subject to position effect variegation (PEV) in conjunction with a lower transgenesis rate, alternative methods such as bacterial (BAC) or yeast (YAC) artificial chromosomes may be a suitable alternative to transposons. BAC and YAC transgenics both exhibit position-independent expression due to the presence of a larger fragment that contain the local surrounding genomic sequence (Giraldo and Montoliu, 2001; Matthaei, 2007), and BAC transgenics are reported to exhibit reproducible expression in 85% of transgenic lines (Gong *et al.*, 2003).

The difficulties encountered with the CRISPR-Cas9 loss-of-function approach are most likely site specific since (i) the introduction of deletions via incorporation of multiple guides typically achieves high mutagenesis efficiencies (Meyer *et al.*, 2015; Sanjana *et al.*, 2016; Zhou *et al.*, 2014) and (ii) the JCSMR Transgenesis Facility report that 50% of NHEJ-dependant projects are successful after the first microinjection session (Burgio laboratory, personal communication). The presence of only a few NGG PAM sites within the NCE itself present difficulties, especially if the guides sequences flanking these sites cannot sufficiently target Cas9 to this region. An alternative CRISPR endonuclease is therefore required. Cpf1 has recently been shown to efficiently cleave target DNA in a manner similar to Cas9, but it targets T-rich motifs and does not require tracrRNA in contrast to Cas9 which requires a G-rich PAM site and the presence of a tracrRNA bound to crRNA to guide the protein to a target sequence (Zetsche *et al.*, 2015). As the *Zic2* NCE was found to be enriched for A and T nucleotides (Chapter 6), the use of Cpf1 in future CRISPR mutagenesis experiments would provide multiple new PAM recognition sequences

within the conserved domain and may prove more successful than Cas9. Cpf1 would also aid in the introduction of the HPE-associated SNVs into the NCE, which would be difficult to achieve with Cas9 unless unusually long homology oligonucleotides are used.

7.3.2 The *ZIC2* NCE likely controls *ZIC2* gene expression

Based on the results shown in this Chapter and Chapters 4, 5 and 6 of this thesis, it can be concluded that the wildtype *ZIC2* NCE possess two main roles: transcription repression and promotion of mRNA decay. In its role as an ER element, the NCE acts to repress transcription of a reporter gene in a node-like signalling environment, with the six HPE-associated SNVs disrupting this ability. Moreover, the addition of TFs such as FOXA2, FOXJ1 and ZIC2 modulate this repressive function in a context-specific manner. Deletion of this region from the murine genome appears to result in an increase in *Zic2* expression, however more data is required to confirm this result. In its role regulating transcript decay, an ARE within the NCE, along with the use of the most distal poly(A) site during gastrulation, determines *ZIC2* half-life via predicted interactions with RBPs and miRNAs. The production of a lncRNA from within the NCE, predicted to be eRNA, suggests a third, as yet unknown role for the *ZIC2* 3'UTR. From these results, I conclude that the *Zic2* NCE likely controls *Zic2* expression during gastrulation.

The introduction of M5 into the NCE is the most likely to disrupt ER function by altering interactions with TFs in the node, whilst M3 and M4 are predicted to disrupt the stability of the *Zic2* transcript due to their location within the putative ARE and nearby the poly(A) site. These three SNVs are therefore the best candidates for CRISPR-mediated HDR mutagenesis in the *Zic2* NCE and can be used to determine if, and how, a single nucleotide change can lead to HPE. Additionally, it would be prudent to delete minimal sections of the NCE to determine the exact function of each region. It is expected that certain regions of the NCE would produce a more prominent phenotype based off the importance of the TFs binding to that region, or the presence of RNA regulatory elements such as the ARE or poly(A) sites. By determining the most active regions of the NCE involved in *Zic2* regulation, or embryo development as a whole, it will allow human geneticists to categorise those SNVs that are most likely to lead to HPE based on their location in the 3'UTR.

The bigene nature of *ZIC2* and *ZIC5* suggests that the NCE may also play a role in regulating *ZIC5* during gastrulation. This hypothesis is supported by the identification of a human proband with *ZIC2*-associated HPE who also exhibits hydrocephaly (Weiss *et al.*, 2017), a known phenotype of *Zic5* null alleles in mice (Inoue et al. 2004; Arkell lab, unpublished data). Future analyses of the NCE CRISPR mutants should include examination of *Zic5* expression and determination of whether the mutant embryos exhibit hydrocephalus.

7.3.3 A 'Goldilocks zone' of *ZIC2* expression may be required to prevent HPE

Analysis of *ZIC2* coding domain mutations has led to a clear expectation that HPE pathogenesis arises from *ZIC2* loss-of-function. It was therefore hypothesised that the NCE in the *ZIC2* 3'UTR would promote *ZIC2* transcription, and that introduction of the six HPE-associated SNVs would result in loss of enhancer function and reduced levels of *ZIC2* transcript. The analysis presented in this thesis, however, suggests the inverse is occurring. The default role of the *ZIC2* 3'UTR appears to be as a repressor of reporter expression in a node-like signalling environment, and the introduction of the six HPE-associated SNVs results in loss of this repressive ability. As such, it is expected that if *ZIC2* is a target of the NCE, the mutations would result in an increase in *ZIC2* expression. Similarly, the proximity of the HPE-associated SNVs to the PASs within the *ZIC2* 3'UTR predicts increased transcript half-life and elevated levels of ZIC2 protein when they are introduced.

The dual roles of the *ZIC2* 3'UTR characterized in this thesis suggest that the 3'UTR acts to reduce *ZIC2* levels during normal development and that mutation of the 3'UTR results in an increase in *ZIC2*. Gain of ZIC2 function has only been implicated in one human case of *ZIC2*-associated HPE thus far, with the CDS mutation Q36P resulting in a 170% increase in ZIC2 transactivation activity on the ApoE promoter when tested *in vitro* (Brown *et al.*, 2005). Recent experiments in *Xenopus* embryos provide further support for this mechanism by demonstrating that both knocking out *zic2* (via morpholino injections) and upregulating *zic2* (via mRNA injections) resulted in the same phenotype: the loss of *foxa2* organizer expression and defects in embryo development (Houtmeyers *et al.*, 2016). It is therefore feasible that mutation of the *ZIC2* 3'UTR leads to elevated *ZIC2* transcript and HPE. Potentially, a 'Goldilocks zone' of *ZIC2* expression is required for normal embryo development (Figure 7.12). If *ZIC2* expression is too little, or too much, defects such as HPE will occur.

It is possible that hypermorphic mutations occur more frequently than thought, but are obscured by embryonic lethality. This correlates with known features of *Zic2*-associated HPE, such as the positive correlation between ZIC2 mutation and HPE severity or lethality (Table 1.2) (Solomon *et al.*, 2010b). Without sufficient evidence from mutant embryos, it is difficult to confirm the existence of this 'Goldilocks zone' directly. Therefore, to confirm if any of the NCE SNVs result in an increase in *ZIC2* expression as predicted in this thesis, the full range of CRISPR mutants attempted in this chapter (tiled NCE deletions), and mutants for the six individual SNVs, need to be made and analysed. Examination of *Zic2* expression via WMISH and RT-qPCR will determine whether *Zic2* is upregulated by the mutations, and phenotyping of embryos will determine if this is sufficient to cause HPE. If the 'Goldilocks zone' hypothesis proposed here proves correct, it will cement a new aetiology for *ZIC2*-associated HPE.

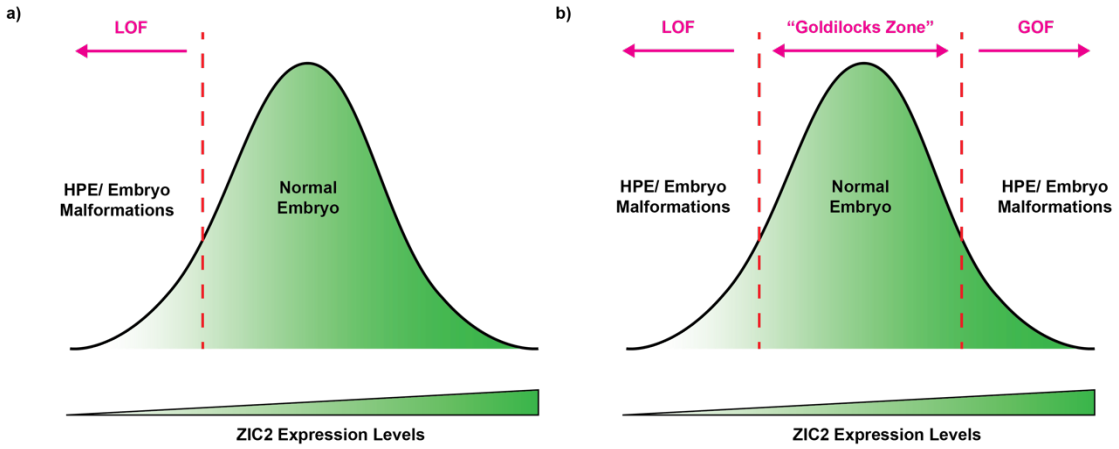


Figure 7.12: The 'Goldilocks zone' hypothesis of *ZIC2*-associated HPE. (a) The majority of *ZIC2*-associated HPE cases can be attributed to loss of *ZIC2* transcript or function (LOF). **(b)** Based on the data presented in this thesis, and evidence from *Xenopus* embryos and one HPE proband, we hypothesis that gain of *ZIC2* transcript or function (GOF) can also lead to embryonic malformations such as HPE. This would result in a 'Goldilocks zone', where *ZIC2* is expression is not too little or too much, but the right level to promote normal embryonic development.

Appendix A1

Table A1.1: Constructs used throughout this thesis.

Construct name	Description	Reference
Gateway Cloning		
pENTR1A, 2B and 3C	Entry vectors used for Gateway Cloning (Invitrogen). Contain multiple cloning site (MCS) surrounding the negative selection gene *ccdB*, all of which is bordered by two *att*L sites for LR recombination based cloning. Each plasmid (1A, 2B and 3C) has a different frame.	Thermo Fisher Scientific
pcDNA3.1/nV5-DEST	pcDNA3.1/nV5-DEST (Invitrogen) Destination vector used for Gateway Cloning. Contains the cytomegalovirus (CMV) promoter upstream of a V5 epitope tag followed by two *att*R sites surrounding a *ccdB* gene for Gateway LR recombination based cloning. In mammalian cells, expresses a V5 tag fused to N- terminus of sequence inserted between *att*R sites.	Thermo Fisher Scientific
pENTR3C-*Foxa2*	See section 2.1	
pENTR1A-*Foxj1*	See section 2.1	
pENTR2B-*NOGGIN*	See section 2.1	
pENTR3C-*ZIC2*	See section 2.1	
pCR8-GW-TOPO-NCE (Wildtype and mutant)	pCR8-GW-TOPO Gateway entry vector containing 608 bp of the wildtype and mutant human *ZIC2* 3'UTR, including the NCE region.	Muenke laboratory (NHGRI, NIH; Maryland, USA)
pGL4.20-DEST-*luc*	See section 2.1	

pGL-β-globin-DEST-*luc*	See section 2.1	
pGL-Hsp68-DEST-*luc*	See section 2.1	
pGL-pZIC3-DEST-*luc*	See section 2.1	
Expression constructs		
pALK4-HA	Constitutively active form of Rat *Alk4* with HA epitope tag in pcDNA3.0	(Bernard *et al.*, 2006)
pV5-*Foxa2*	See section 2.1	
pV5-*Foxj1*	See section 2.1	
pV5-*NOGGIN*	See section 2.1	
pV5-*ZIC2*	See section 2.1	
pV5-*ZIC2*-3'UTR	See section 2.1	
Luciferase reporter constructs (non-Gateway Destination)		
pGL4.20	pGL4.20(luc2/Puro) reporter vector. Contains a MCS upstream of luciferase reporter gene *luc2* (codon optimised for expression in mammalian cell lines). The backbone has been engineered to reduce undesired non-specific transcription factor binding.	Promega
pGL4.20-β-globin-NCE (wildtype)	See section 2.1	
pGL4.20-Hsp68-NCE (wildtype)	See section 2.1	
pGL4.20-pZIC3-NCE (wildtype and mutant)	See section 2.1	

pGL-$(CAGA)_{12}$-*luc*	Twelve tandem optimised CAGA boxes (a SMAD3 and SMAD4 binding site), upstream of a TK minimal promoter followed by the luciferase reporter gene. Used to measure levels of Nodal and Activin signalling activation.	(Dennler *et al.*, 1998; Kumar *et al.*, 2001)
pGL-β-globin-TOPFLASH-*luc*	Three tandem optimised TCF binding sites upstream of a β-globin minimal promoter (isolated from pKS:β-globin:*lacZ*) followed by the luciferase reporter gene. Used to measure levels of WNT/β-catenin signalling activation.	Helen Bellchambers (Arkell laboratory)
pGL-BRE-*luc*	Three distinct BMP response elements (BREs), upstream of a TK minimal promoter followed by the luciferase reporter gene. Used to measure levels of BMP signalling activation.	(Korchynskyi and Ten Dijke, 2002; Morikawa *et al.*, 2011)
pGl-pZIC3-*luc*	A 340 bp untested ZIC3 promoter upstream of a luciferase reporter (pGL3.0)	Muenke laboratory (NHGRI, NIH; Maryland, USA)
Riboprobes		
pCDNA-*CDX2*	The complete CDS of *Cdx2* (Accn. No: NM007673) in pCDNA 3.1	Stefan Broer, ANU
pGEM-T-Easy-*Dand5*	See section 2.1	
pGEM-T-Easy-*Dkk1*	The complete *Dkk1* CDS (Accn. No. JN966751) in pGEM-T-Easy	(Ip *et al.*, 2014)
pBS-*Foxa2*	1.5 kb complete murine *Foxa2* CDS (clone c21) in pBS-KS II	(Sasaki and Hogan, 1993)
pYX-ASC-*Foxj1*	Complete murine *Foxj1* CDS (Accn. No: BC082543) in pYX-ASC	(Cruz *et al.*, 2010)
pBS-*Lefty1/2*	500 bp mature protein coding region for both murine *Lefty1* and *Lefty2* in pBS-KS II	(Meno *et al.*, 1996)
pBS-*Lhx1*	2.4 kb fragment of the murine *Lhx1* (*Lim1*) CDS in pBS-KS II	(Shawlot and Behringer, 1995)
pBS-*Nodal*	2 kb complete murine *Nodal* CDS in pBS-KS II	(Conlon *et al.*, 1994)
pGEM-T-Easy-*Noto*	See section 2.1	

pT7T3D-*Nppa*	1 kb fragment of *Nppa* CDS, IMAGE 402095 (Accn. No: W77688) in pT7T3D	Dominic Norris, MRC Harwell, UK
pBS-*Pitx2*	Complete 1754 bp *Pitx2* gene (IMAGE 371701) in pBS-KS II	(Ryan *et al.*, 1998)
pBS-*Pkd1l1*	805 bp fragment of the *Pkd1l1* CDS in pBS-KS II	(Field *et al.*, 2011)
pGEM-T-Easy-*Rfx3*	See section 2.1	
pBS-*Shh*	2.6 kb complete *Shh* CDS in pBS-KS II	(Echelard *et al.*, 1993)
pBS-*Sox2*	530 bp fragment of the *Sox2* CDS in pBS-KS II	(Wood and Episkopou, 1999)
pBS-*Sox3*	900 bp fragment of the *Sox2* CDS in pBS-KS II	(Wood and Episkopou, 1999)
pT7T3D-*Sox17*	1111 bp of the *Sox17* CDS, IMAGE 1529001 (Accn No: AW985818) in pT7T3D	Dominic Norris, MRC Harwell, UK
pGEM-T-Easy-*Zic2*	738 bp of *Zic2* Exon 1 to beginning of zinc finger domain (Accn. No. NM009574.1) in pGEM-T-Easy	(Elms *et al.*, 2003)
pGEM-T-Easy-*Zic3*	Bases 646-1266 of *Zic3* (Accn. No. NM009575) in pGEM-T-Easy. Exon 1 up to zinc finger domain	(Elms *et al.*, 2004)
pGEM-T-Easy-*Zic5*	788 bp fragment spaning the 5' UTR and exon1 of *Zic5* in pGEM-T-Easy	Arkell laboratory
Other		
pGEM-3'UTR-WT	See section 2.1	
pBB232	Contains the Pbase (*PiggyBac*) transposase open reading frame.	Bryan Bjork (Harvard Medical School, Boston, MA; Bjork et al. 2010)

pBB256, pBB259 and pBB262	Constructs for the delivery of desired transgene cargo. Each construct contains two inverted terminal repeats. The MCS is flanked by the repeats, along with two (pBB259) or four (pBB262) copies of the cHS4 chicken insulator fragment. For more information, see Chapter 7.	Bryan Bjork (Harvard Medical School, Boston, MA; Bjork et al. 2010)
p1229	The 4 kb *lacZ* gene and SV40 PolyA tail in pBS-KS II with a β-globin minimal promoter in pBS-KS II	Nick Warr (Arkell laboratory)
pBB262-*ZIC2* NCE (wildtype and mutant)	See section 2.1	
pBB262-NCE-*lacZ* (wildtype and mutant)	See section 2.1	
pBB262-*lacZ*	See section 2.1	
pmNETg	The 1617 bp mNet fragment cloned upstream of lacZ in p1229	Ruth Arkell (Arkell laboratory)
pBB262-mNet	See section 2.1	
pBB262-mNet-*lacZ*	See section 2.1	
pKS:β-globin:*lacZ*	The 52 bp β-globin minimal promoter upstream of a *lacZ* reporter in pBS-KS II	(Yee and Rigby, 1993)
pBS-Hsp68	The 293 bp heat shock protein 68 (Hsp68) promoter in pBS-KS II	(Adachi *et al.*, 1999)
pGEM-T-Easy-Hsp68	See section 2.1	
pTracer-CMV-Bsd-*lacZ*	Two Tracer™ mammalian expression vector with CMV promoter that express GFP fused to the selectable marker Blasticidin. Contains a *lacZ* reporter.	Tscharke laboratory (ANU), Thermo Fisher Scientific
pTOPO-Blunt-II	The pTOPO-Blunt-II vector is designed to clone blunt-ended PCR products generated by thermostable proofreading polymerases. Used to efficiently clone in CRISPR guide strands.	Thermo Fisher Scientific

pTOPO-Blunt-II-G2	*ZIC2* NCE guide 2 (G2) cloned into the pTOPO-Blunt-II vector.	JCSMR Transgenesis Facility
pTOPO-Blunt-II-G2	*ZIC2* NCE guide 3 (G3) cloned into the pTOPO-Blunt-II vector.	JCSMR Transgenesis Facility

Table A1.2: Oligonucleotides used throughout this thesis.

Oligonucleotide Name	Sequence (5' – 3')	For
Cloning and sequencing		
RA104	TAC GAC TCA CTA TAG GG	Sequencing plasmids (T7)
RA105	ATT AAC CCT CAC TAA AG	Sequencing plasmids (T3)
RA561	GTT TTC CCA GTC ACG AC	Sequencing plasmids (M13)
RA562	ACA GGA AAC AGC TAT GAC	Sequencing plasmids (M13)
RA1411	AAC TAA GCA GAA GGC CA	Sequencing plasmids (pENTR)
RA1412	GTG CAA TGT AAC ATC AG	Sequencing plasmids (pENTR)
RA1443	TAG CTG TAC ATC TAG ATG TGT ACA TAG CGG ACT CCT CCT TTC	Cloning the *ZIC2* NCE into pBB262-*ZIC2* NCE via In-Fusion Cloning. Contains a *Xba*I restriction site.
RA1445	TTG TCA ATC CTC AGC TG	Sequencing the *ZIC2* 3'UTR
RA1446	CCT AAA GTG ATG GGC TT	Sequencing the *ZIC2* 3'UTR
RA1464	TAG GGG GCC CTC TAG AGC GGC CGC TGG GTC GAA TTC GCC CTT GTC AAT	Cloning the *ZIC2* NCE into pBB262-*ZIC2* NCE via In-Fusion Cloning. Contains *Xba*I and *Not*I restriction sites.
RA1506	TCAGATCTCGAACTAGTGGATCCCC	Cloning the β-globin minimal promoter into pGL-β-globin-DEST-*luc*. Contains a *Bgl*II restriction site.
RA1507	CTAAGCTTTCGGCTAGAAGCAAATG	Cloning the β-globin minimal promoter into pGL-β-globin-DEST-*luc*. Contains a *Hind*III restriction site.
RA1545	ATGGTACCCAGCCTGTGGATGTTCA	Cloning *Foxj1* into pENTR1A-*Foxj1.* Contains a *Knp*I restriction site.
RA1546	ATGCGGCCGCCCTGACTTGAGCACTGT	Cloning *Foxj1* into pENTR1A-*Foxj1.* Contains a *Not*I restriction site.

RA1570	GGAAACCCTAGTCAGAA	Cloning *Dand5* into pGEM-T-Easy-*Dand5.*
RA1571	CATGAATGAACCTTGGC	Cloning *Dand5* into pGEM-T-Easy-*Dand5.*
RA1590	ATTGAGCTCCTTGCACA	Cloning *Noto* into pGEM-T-Easy-*Noto.*
RA1591	TGGTGAGTGACATTCAG	Cloning *Noto* into pGEM-T-Easy-*Noto.*
RA1592	AGCGAAGTCGAAAGTGA	Cloning *Rfx3* into pGEM-T-Easy-*Rfx3.*
RA1593	TCCTTCTTGACACCTGA	Cloning *Rfx3* into pGEM-T-Easy-*Rfx3.*
RA1596	GACTCGAGTGAATGGTACGTGTGAC	Cloning the *ZIC2* 3'UTR into pGEM-3'UTR-WT. Contains an *Xho*I restriction site.
RA1597	GACTCGAGTTTGCAGTTTCAACACATTT	Cloning the *ZIC2* 3'UTR into pGEM-3'UTR-WT. Contains an *Xho*I restriction site.
RA1626	TCAAGCTTGGATTTGTGGGGTTTCG	Cloning the Hsp68 promoter into pGEM-T-Easy-Hsp68. Contains an *Hind*III restriction site.
RA1640	AGATCTGATTACGCCAAGCTTGC	Cloning the Hsp68 promoter into pGEM-T-Easy-Hsp68. Contains an *Bgl*II restriction site.
RA1776	ATGGTACCTCAGAGGCATGGAGCGCT	Cloning *NOGGIN* into pENTR2B-*NOGGIN.* Contains an *Kpn*I restriction site.
RA1777	TAGCGGCCGCAGTCGGTGGGGATCGATC	Cloning *NOGGIN* into pENTR2B-*NOGGIN.* Contains an *Not*I restriction site.
RA1824	CAGTCGACTACTTCCAGCATGCTGGGA	Cloning *Foxa2* into pENTR3C-*Foxa2.* Contains a *Sal*I restriction site.
RA1825	GACTCGAGCATGTCCCCAGTGACCAGA	Cloning *Foxa2* into pENTR3C-*Foxa2.* Contains a *Xho*I restriction site.
RT-qPCR		
RA96	ACA GGT TGG AGC TGC TTT GT	RT-qPCR of mouse *Zic2*

RA99	TCT CTG TAT GTC TGA AAC TG	RT-qPCR of mouse *Zic2* 3'UTR PolyA site mP1, Amplification of the *Zic2* 3'UTR for 3'RACE
RA252	TCG TTG CGG AAG CAC ATG AA	RT-qPCR of mouse *Zic2*
RA1177	GTC AAC CAC ATC CGC GTG CA	RT-qPCR of human *ZIC2*
RA1247	GGT CGC AGC CCT CAA ACT CAC	RT-qPCR of human *ZIC2*
RA1604	CTT CAG CGG AGG ACT CTA CG	RT-qPCR of human *PPP1R8*
RA1605	GGG GCA AGG TTT GGG TAT GG	RT-qPCR of human *PPP1R8*
RA1606	GGAATGCAAAAAGAGTTGTCCTG	RT-qPCR of human *ABCE1*
RA1607	CGAGGGATAGGCAACCTGTG	RT-qPCR of human *ABCE1*
RA1719	GCG CAG CCA TCC TGG AGT A	RT-qPCR of mouse *H2afz*
RA1720	CCG ATC AGC GAT TTG TGG A	RT-qPCR of mouse *H2afz*
RA1766	CGA CTG GAG CAG CTA CTA TGC	RT-qPCR of human *FOXA2*
RA1767	TAC GTG TTC ATG CCG TTC AT	RT-qPCR of human *FOXA2*
RA1768	CAA CTT CTG CTA CTT CCG CC	RT-qPCR of human *FOXJ1*
RA1769	CGA GGC ACT TTG ATG AAG C	RT-qPCR of human *FOXJ1*
RA1780	CCA ACT TCA ATG AAT GGT AC	RT-qPCR of mouse *Zic2* 3'UTR poly(A) site mP0
RA1781	AGA CGA TCC TAA GGT GC	RT-qPCR of mouse *Zic2* 3'UTR poly(A) site mP0
RA1782	CAT GTG GGA AGG AAA GTG	RT-qPCR of mouse *Zic2* 3'UTR poly(A) site mP1
RA1714	AGT CTC CCT TCT GTT TC	RT-qPCR of mouse and human *ZIC2* 3'UTR poly(A) site mP2
RA1715	CAC ATG TAA ACC CAG CT	RT-qPCR of mouse and human *ZIC2* 3'UTR poly(A) site mP2

RA1716	CTT GCT ACA GAG GAA AC	RT-qPCR of mouse *Zic2* 3'UTR poly(A) site mP3
RA1717	CAA ACC ACT TAA ACC TTC	RT-qPCR of mouse *Zic2* 3'UTR poly(A) site mP3
RA1730	CGT CGA GCC CAG TGT TAC CAC CAA GAA GG	RT-qPCR of mouse *Ubc*
RA1731	CCC CCA TCA CAC CCA AGA ACA AGC ACA AG	RT-qPCR of mouse *Ubc*
Genotyping		
RA247	GGG AGA AACT CTT TCC AGT GTG AG	*Kumba* HRMA and ADA genotyping
RA248	GCT TCT TCC TGT CGC TGC TG	*Kumba* HRMA and ADA genotyping
RA748	CAC ACA CAC ATT TCT CTG TCC	mNet HRMA genotyping (control)
RA749	TAG CTC AGT GCT TGC AAG GTT A	mNet HRMA genotyping (control)
RA1058	ATC GGC CTC AGG AAG ATT	mNet HRMA genotyping, *PiggyBac* pBB262-NCE-*lacZ* genotyping, *PiggyBac* pBB262-mNet-*lacZ* genotyping and *PiggyBac* pBB262-*lacZ* genotyping
RA1059	CAC ATC AGA CTA ATC ACC TCG	mNet HRMA genotyping, *PiggyBac* pBB262-mNet-*lacZ* genotyping
RA1547	GTG GGC ATA AAC TGT TTC AG	*PiggyBac* pBB262-NCE-*lacZ* genotyping
RA1548	GCT AGC TGT ACA TCT AGA AC	*PiggyBac* pBB262-*lacZ* genotyping
RA1778	GGC ATG AAA AGC AAA ATC TCT CT	Genotyping of CRISPR mutant embryos
RA1779	CCT CCA AAG AAT GCA AGG TGG	Genotyping of CRISPR mutant embryos
Other		
RA98	TTT CTC CCA TTC CCT GTT CC	Amplification of the *Zic2* 3'UTR for 3' RACE
RA748	CAC ACA CAC ATT TCT CTG TCC	Genomic DNA detection in mouse RNA samples

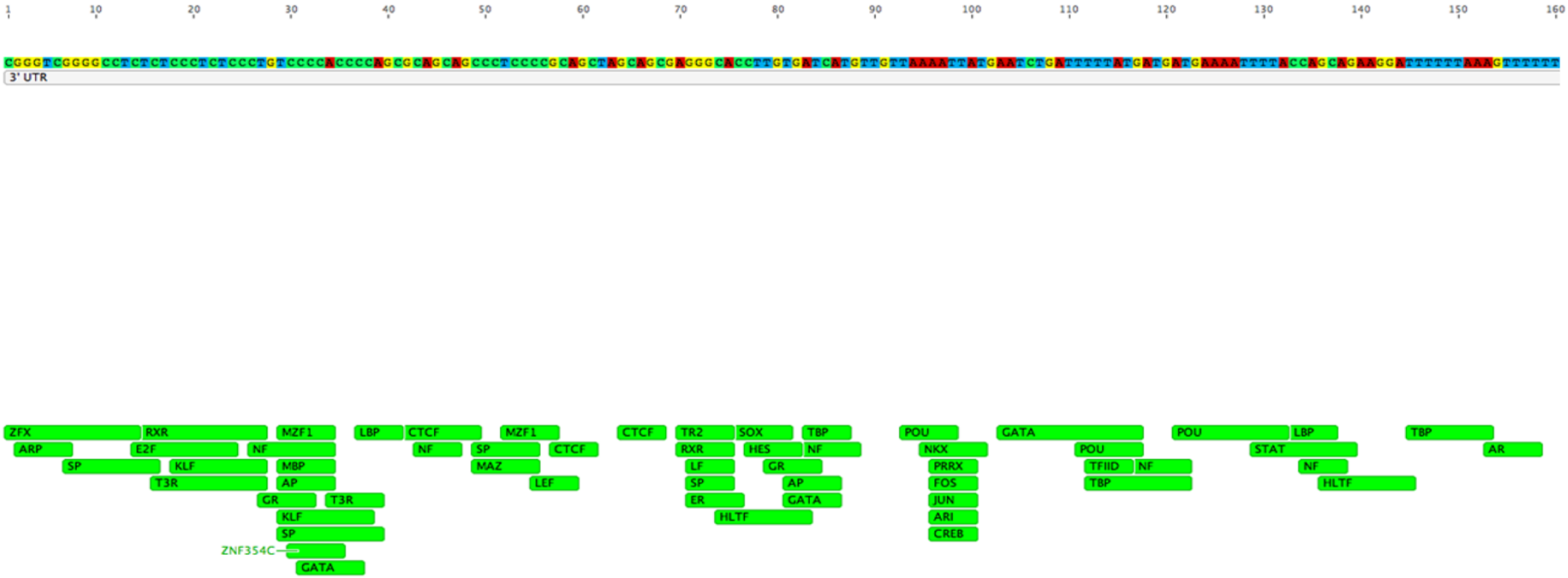
RA749	TAG CTC AGT GCT TGC AAG GTT A	Genomic DNA detection in mouse RNA samples
RA1595	TGA GAT CTT TAA AGG GGC GGT TGA	Genomic DNA detection in human RNA samples
RA1622	GCG TAG TAA ACT TGC AG	Genomic DNA detection in human RNA samples
RA1703	GGC CAC GCG TCG ACT AGT ACT TTT TTT TTT TTT TTT TV	dT-adapter oligonucleotide for 3’ RACE
RA1704	GGC CAC GCG TCG ACT AGT AC	3' RACE PCR oligonucleotide for recognition of RA1703

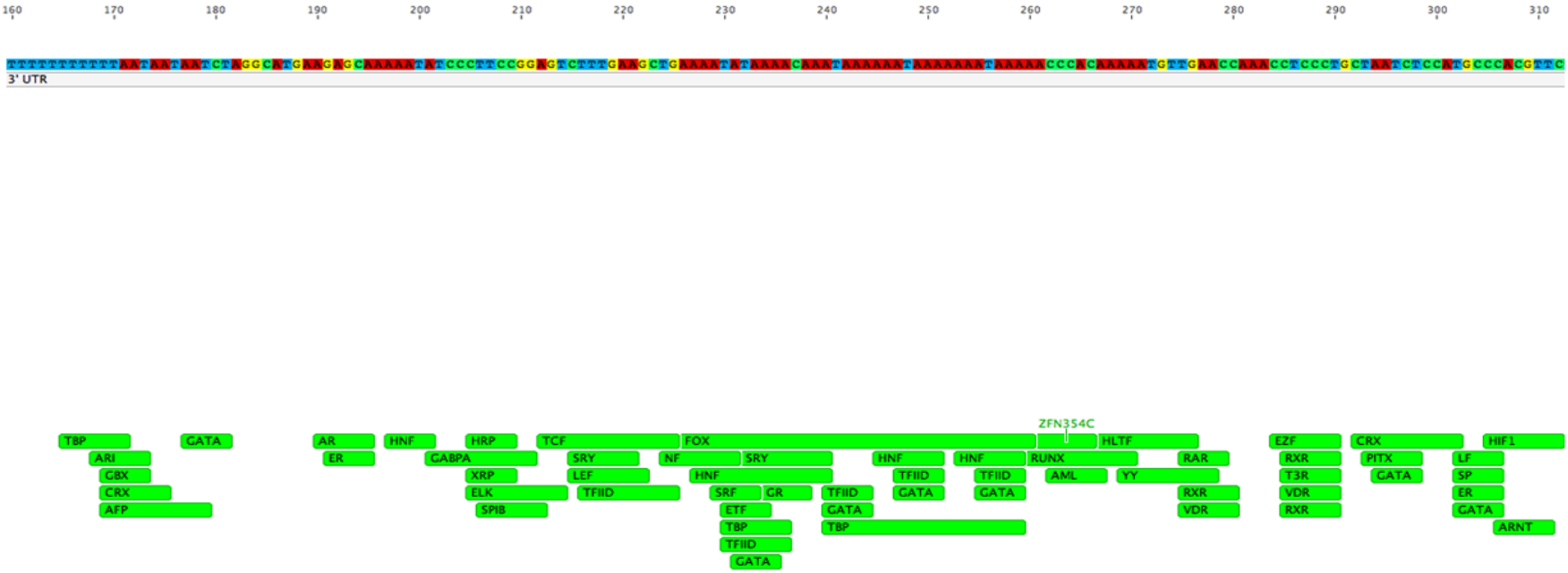
Appendix A2

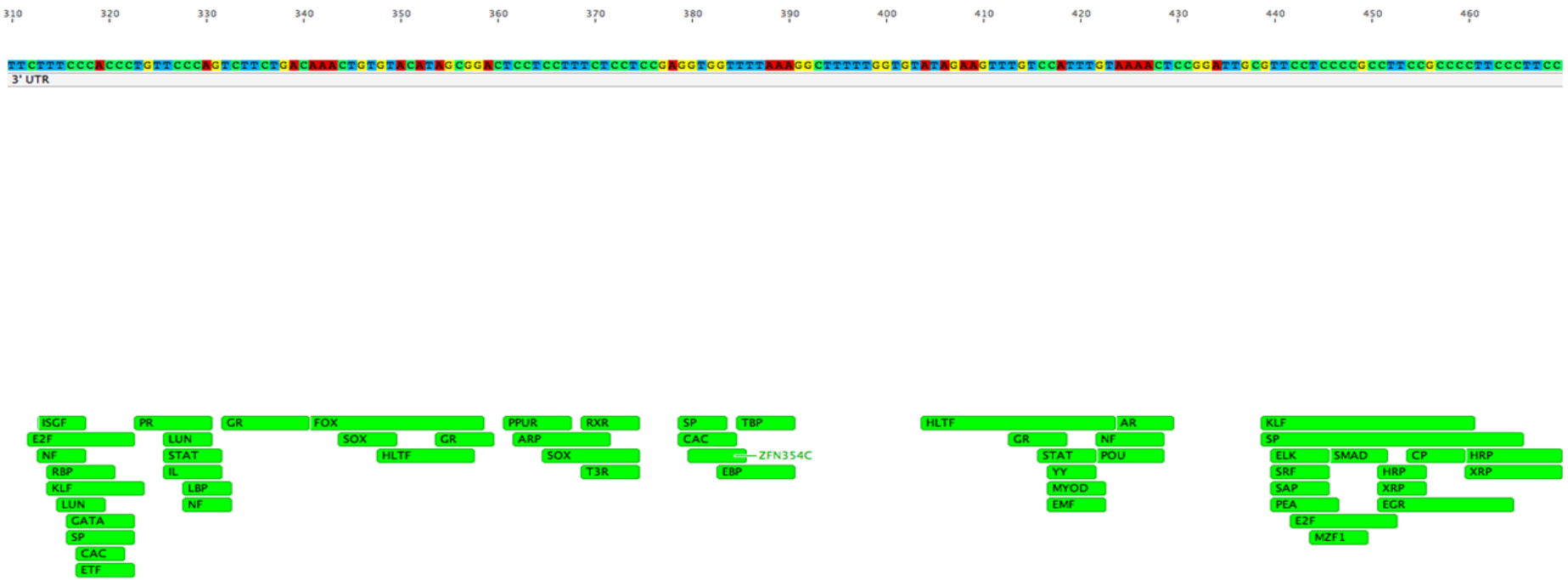
5′ GGACAAACACAAACCCTGTTAATTATAGAATGGACCAAATACATTTTTAAAAGAAAACTGA-
GACCAATCAGATGGAAATGGAGTTTTAAGGCAAGAGGCCATATATAGGGCTACATCTTGTTAATTGCAATTGT
CCAGGAAGGTTTTGGGCAAGATCCAAAAGTAGCCATGCCCTTTTCTCAGGATAGAAAATATGTTTTGGCATTT
GAAGCATTTTTTACAAAATCTTTACACTACTTTTTCTTCCCCTTCCTCTTGCTCTCTGCACACCCCATTCTTAAA
CTCCTCCAATTCATTTTAACACTTGTCCTGTTTCTTGAGAGGAAGTTATAGAAGGCTTGTTGGTGGTGGTGAT
GTTAAACTGATGGAAATTCTTTTTCGCCTTAGTGGTGATTGTTTAAACTCTCACAGTCTTAAACCGTGCCAAA
GTCCTGTTATGTCTTGAACTTTTCCTCAAAGCATTACACTTGTGAATGTATTTTTGTCTAATAGGGTCGAAACT
GTTGTTCAGTATTTTTTCAGGCTGAGGATGTGATGTTACTCTACACATTGTGACGTTTAGTATACAGTTGCCTT
TTGTAATAACTTTTTTTTTGTAAATACATATCCATTGATGCCATATTTATCGTTTGTAATTTAATTATTGCTACAAG
TGCCGGGAACTGAACAATATTTATGGATAAATGTTTTCTAACAAATTCTGTACAGCTTTTGATTATAACTGCTTT
AGCATTAAAAATTGTTTTGAAAGAAGAACACAATTTACAATTTTGGAACCACTGACTCCTTTCTCTTGTTTTG
TAACAGCCTTCTTCTACAAAGAGGAGATGTGAGCAAATTAAATCTTGTTTGTTGGTATTTATAACTCACTCAGA
TCCCTTTTTTAATTGTTAAATTATTTTTCTATTACAGTATAAATTCCTTACAGTGTCAGTTTCCATCTGGGAAGAC
TCTCCTTTCTTTATCTCTATCTCAGATGGTTGTTTAACTGCGAGTTTAAATGTGTTTGTCCTGGATTTTCGGCAT
GCAAATCAAATATTACTGATCAATTCAGTTAGTGGCCATGACATCTCAATCTTGTACTTCAAAGACTGAGAAG
CTGGATTTAATCATCCCTGCCCTACATATATAAACATAAGGTAACCTACTGAATTTTATGTCCCTTAGTTCTTTAT
TACCTTACATAAAAATGAAAATTGCGGCAGGATGCATGTCTGTCTGTTCTATCTAGAGATCACCCATATACCTAT
ATATGTTTGTATCTATGACTTATCTAATCTGCCTATCAATCTATCTAGTAGCTATCTATATATTTTCAAAAGATAGC
TTATGTCTAAAACAGTGGTGATGAGTAAGGCCAGTTGAGCATTGCTTACTTATGGTTAAAGTGCTTCTTAAAA
GAACCATAGTCCATTTACAATTTTGGAAGGCAAAGGCTGATTTGTTTGCTGTATATAGTTCAATTCATAATTATC
GCAATTATCCATAACATTTATATAGCGGTGTAATAACTGCAGCAGTTCTGAAATGATGCTATGGGAAAAAAATT
GCAAAATATGTATTTTTAAAGTTCATGATTTGTAGGCAAAGATGTTAAGAGCATTGTTTCCATTAAAATCAATA
ACAATGAAAAACGGTTGTTTGCTTTGTGATAAAATGTTAACAATCCATTTAATCTTCAGTGAAGCAACTCATTT
GGACAAACAGTTCTTTTACAGTGGTTATATGGAAAAGAAATTGATTGCTATTTTGGGGGGCTGGAGTGGGTA
GAGTATTGAGACCTTTTTTTATTAGGGTGCTGTTTGTTATTGAGAGCCCAATCTCTGCATGAATAACAGAAAGG
TGGACATAAGGAATTGTAAAGTATTTAGAATGTATTGAATAGGCTTAAGTACCTCCTTTAAGGGGCAATGCTCT
AGGTTTTTGGTGGCAGTCAATTTGGTATTATATATGATCATTTTCAATTCTAGAATTTTGTTTATTGTTCTTTTTG
AAGAAATAAAGTCTTGGCACATCTTATTTATGTTATAA 3′

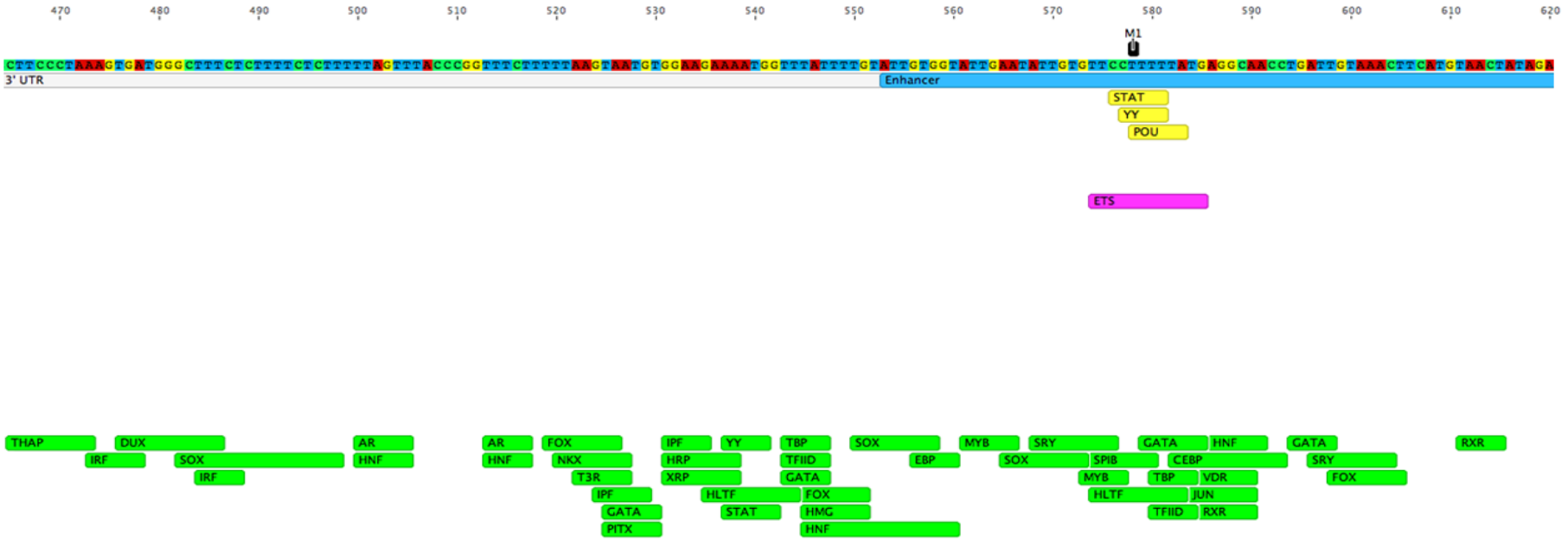
Figure A2.1: Sequence of the *ZIC3* 3'UTR and putative NCE region. A putative *ZIC3* NCE (blue) was identified within the *ZIC3* 3'UTR by E. Roessler and M. Muenke (personal communication, unpublished, NIH) in the same bioinformatics screen that identified the *ZIC2* NCE in the *ZIC2* 3'UTR.

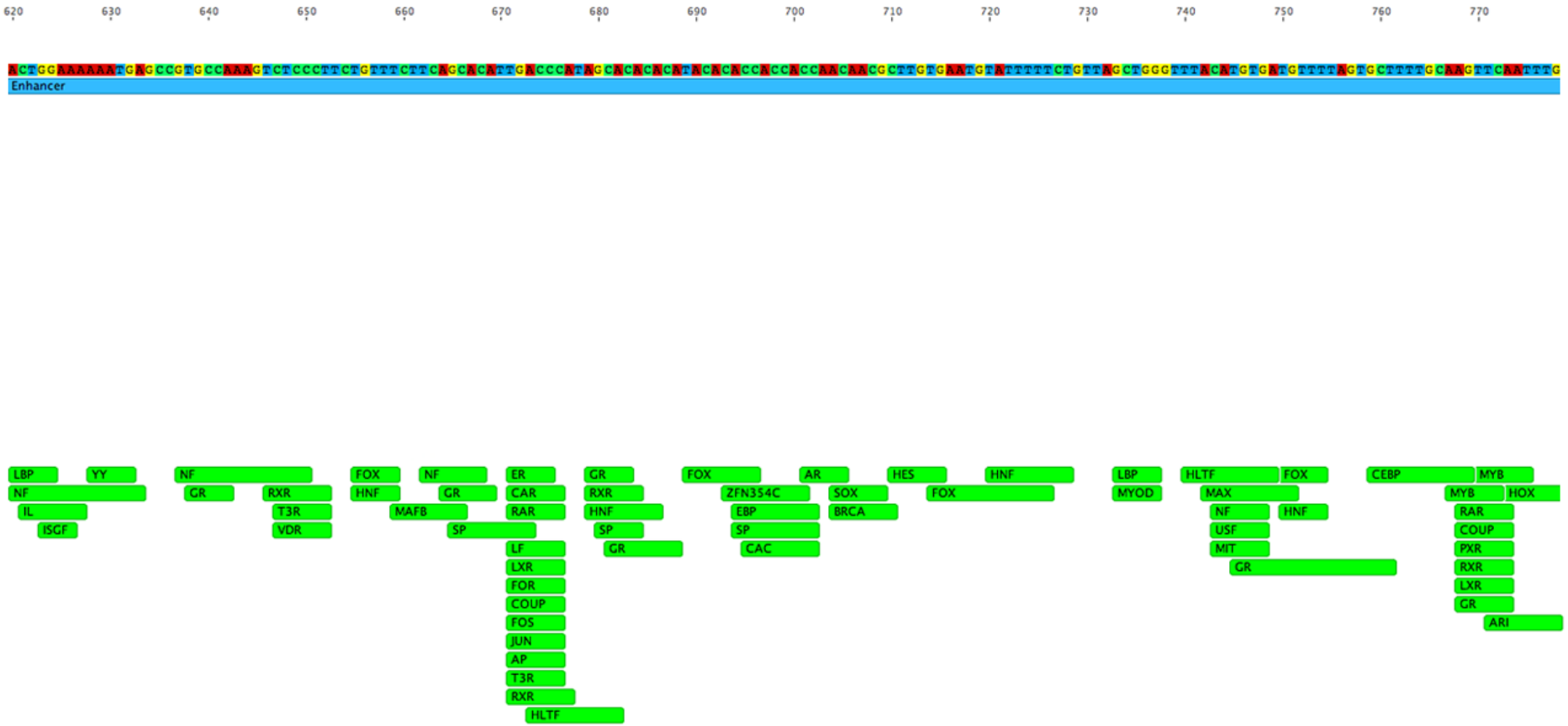
Figure A2.2: Predicted transcription factor binding sites in the *ZIC2* 3'UTR. Pink, mutations that disappear when the corresponding SNVs are introduced; yellow, binding sites created when the corresponding SNVs are introduced; and green, binding sites that remain unchanged in both the wildtype and mutated 3'UTR. Binding sites were predicted by ENCODE, JASPAR and TRANSFAC (Mathelier *et al.*, 2014; Matys *et al.*, 2006; Rosenbloom *et al.*, 2013). HPE-associated SNVs are annotated in black.

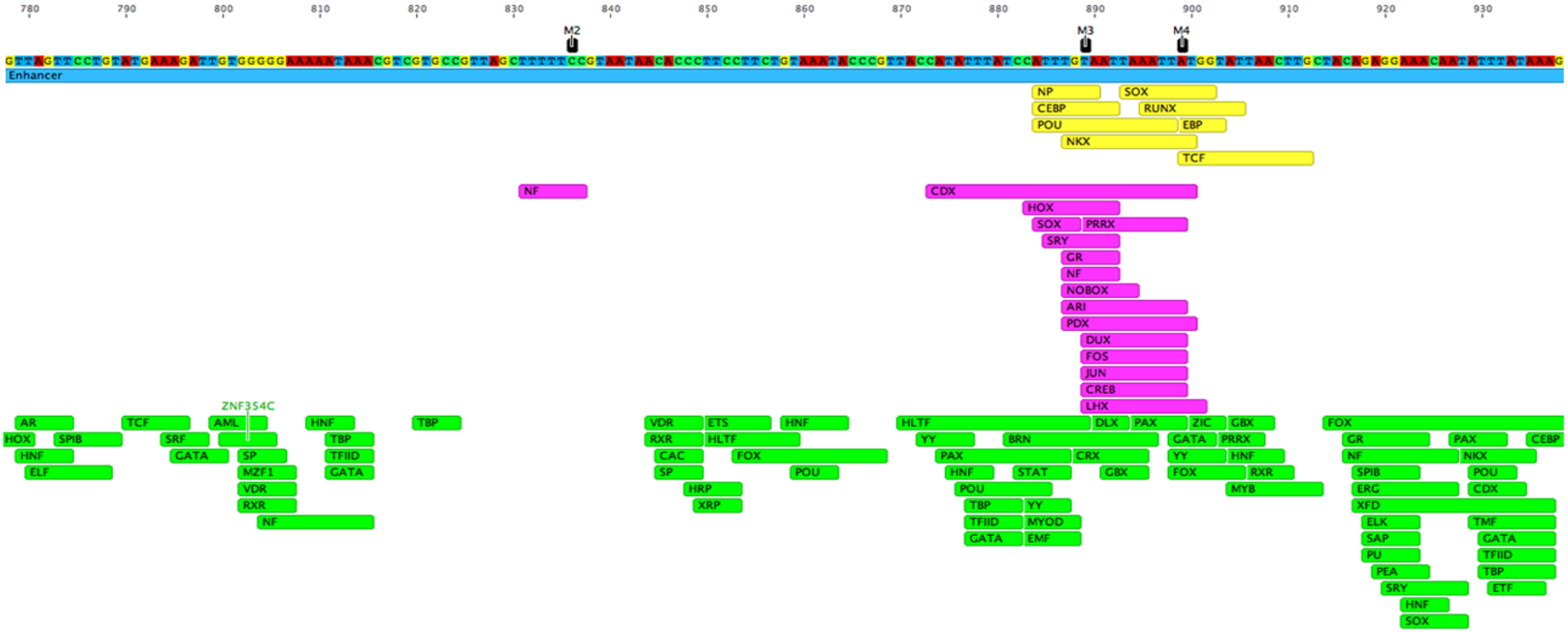

1 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160
CGGGTCGGGGCCTCTCTCCCTCTCCCTGTCCCCACCCCAGCGCAGCAGCCCTCCCCGCAGCTAGCAGCGAGGGCACCTTGTGATCATGTTGTTAAAATTATGAATCTGATTTTTATGATGATGAAAATTTTACCAGCAGAAGGATTTTTTAAAGTTTTTT
3' UTR
ZFX
RXR
MZF1
ARP
E2F
NF
SP
KLF
MBP
T3R
AP
GR
T3R
KLF
SP
ZNF354C
GATA
LBP
CTCF
MZF1
NF
SP
CTCF
MAZ
LEF
CTCF
TR2
SOX
TBP
RXR
HES
NF
LF
GR
SP
AP
ER
GATA
HLTF
POU
NKX
PRRX
FOS
JUN
ARI
CREB
GATA
POU
TFIID
NF
TBP
POU
LBP
STAT
NF
HLTF
TBP
AR

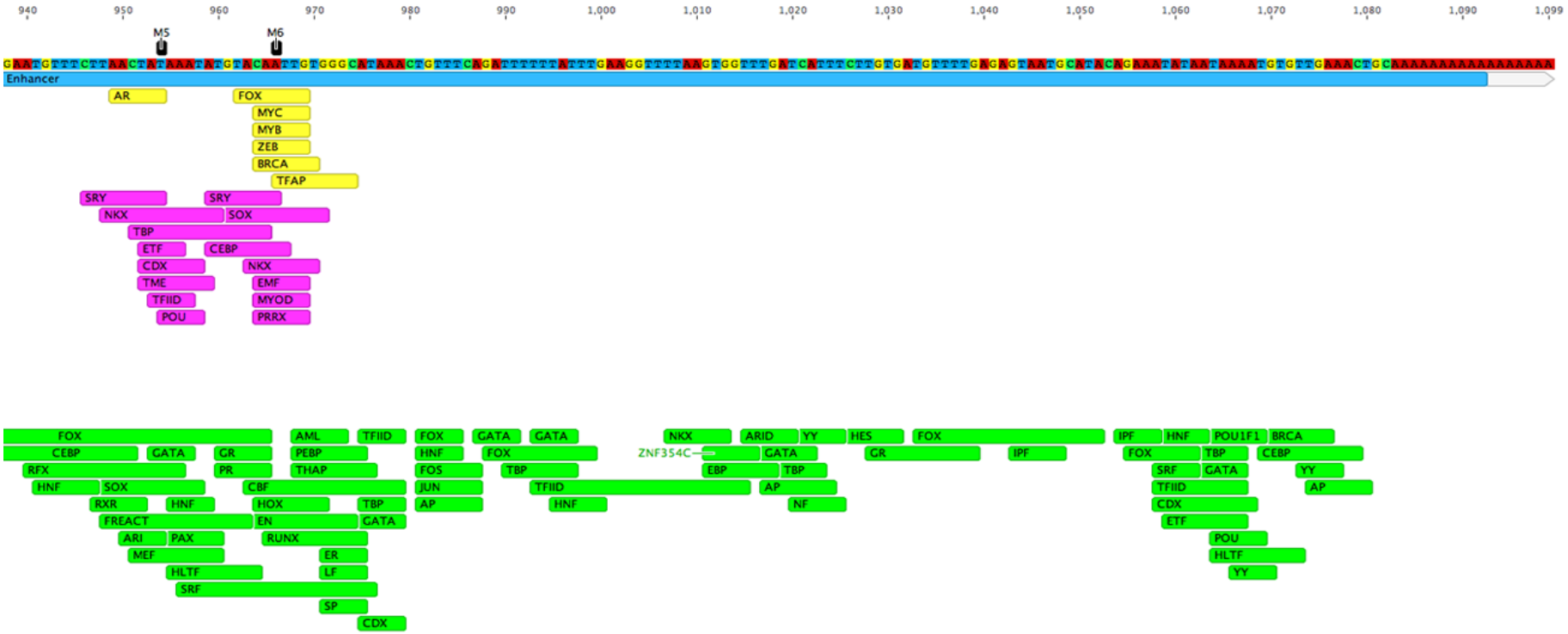

160
170
180
190
200
210
220
230
240
250
260
270
280
290
300
310
3' UTR
ZFN354C
TBP
GATA
ARI
GBX
CRX
AFP
AR
ER
HNF
GABPA
HRP
XRP
ELK
SPIB
TCF
LEF
TFIID
FOX
SRY
NF
SRY
HNF
TFIID
SRF
GR
ETF
TBP
TFIID
GATA
HLTF
HNF
TFIID
GATA
HNF
TFIID
GATA
TBP
RUNX
AML
YY
RAR
RXR
VDR
EZF
RXR
T3R
VDR
RXR
CRX
PITX
GATA
HIF1
LF
SP
ER
GATA
ARNT


310
320
330
340
350
360
370
380
390
400
410
420
430
440
450
460
TTCTTTCCCACCCTGTTCCCAGTCTTCTGACAAACTGTGTACATAGCGGACTCCTCCTTTCTCCTCCGAGGTGGTTTTAAAGGCTTTTTGGTGTATAGAAGTTTGTCCATTTGTAAAACTCCGGATTGCGTTCCTCCCCGCCTTCCGCCCCTTCCCTTCC
3' UTR
ISGF
PR
GR
FOX
PPUR
RXR
SP
TBP
HLTF
AR
KLF
E2F
LUN
SOX
GR
ARP
CAC
GR
NF
SP
NF
STAT
HLTF
SOX
ZFN354C
STAT
POU
ELK
SMAD
CP
HRP
RBP
IL
T3R
EBP
YY
SRF
HRP
XRP
KLF
LBP
MYOD
SAP
XRP
LUN
NF
EMF
PEA
EGR
GATA
E2F
SP
MZF1
CAC
ETF


470
480
490
500
510
520
530
540
550
560
570
580
590
600
610
620
M1
3' UTR
Enhancer
STAT
YY
POU
ETS
THAP
DUX
AR
IRF
SOX
HNF
IRF
AR
HNF
FOX
NKX
T3R
IPF
GATA
PITX
IPF
YY
HRP
XRP
HLTF
STAT
TBP
TFIID
GATA
FOX
HMG
HNF
SOX
EBP
MYB
SRY
GATA
HNF
SOX
SPIB
CEBP
MYB
TBP
VDR
HLTF
JUN
TFIID
RXR
GATA
SRY
FOX
RXR


620
630
640
650
660
670
680
690
700
710
720
730
740
750
760
770
ACTGGAAAAAATGAGCCGTGCCAAAGTCTCCCTTCTGTTTCTTCAGCACATTGACCCATAGCACACACATACACACCACCACAACAACGCTTGTGAATGTATTTTTCTGTTAGCTGGGTTTACATGTGATGTTTTAGTGCTTTTGCAAGTTCAATTTG
Enhancer
LBP
YY
NF
NF
GR
RXR
IL
T3R
ISGF
VDR
FOX
NF
ER
HNF
GR
CAR
MAFB
RAR
SP
LF
LXR
FOR
COUP
FOS
JUN
AP
T3R
RXR
HLTF
GR
RXR
HNF
SP
GR
FOX
AR
HES
HNF
ZFN354C
SOX
FOX
EBP
BRCA
SP
CAC
LBP
MYOD
HLTF
FOX
MAX
NF
HNF
USF
MIT
GR
CEBP
MYB
MYB
HOX
RAR
COUP
PXR
RXR
LXR
GR
ARI


780
790
800
810
820
830
840
850
860
870
880
890
900
910
920
930
M2
M3
M4
GTTAGTTCCTGTATGAAAGATTGTGGGGGAAAAATAAACGTCGTGCCGTTAGCTTTTTCCGTAATAACACCCTTCCTTCTGTAAATACCCGTTACCATATTTATCCATTTGTAATTAAATTATGGTATTAACTTGCTACAGAGGAAACAATATTTATAAAG
Enhancer
NP
SOX
CEBP
RUNX
POU
EBP
NKX
TCF
NF
CDX
HOX
SOX
PRRX
SRY
GR
NF
NOBOX
ARI
PDX
DUX
FOS
JUN
CREB
LHX
ZNF354C
AR
TCF
AML
HNF
TBP
HOX
SPIB
SRF
TBP
HNF
GATA
SP
TFIID
ELF
MZF1
GATA
VDR
RXR
NF
VDR
ETS
HNF
RXR
HLTF
CAC
FOX
SP
POU
HRP
XRP
HLTF
DLX
PAX
ZIC
GBX
YY
BRN
GATA
PRRX
PAX
CRX
YY
HNF
HNF
STAT
GBX
FOX
RXR
POU
MYB
TBP
YY
TFIID
MYOD
GATA
EMF
FOX
GR
PAX
CEBP
NF
NKX
SPIB
POU
ERG
CDX
XFD
ELK
TMF
SAP
GATA
PU
TFIID
PEA
TBP
SRY
ETF
HNF
SOX


940
950
960
970
980
990
1,000
1,010
1,020
1,030
1,040
1,050
1,060
1,070
1,080
1,090
1,099
M5
M6
Enhancer
AR
FOX
MYC
MYB
ZEB
BRCA
TFAP
SRY
SRY
NKX
SOX
TBP
ETF
CEBP
CDX
NKX
TME
EMF
TFIID
MYOD
POU
PRRX
FOX
AML
TFIID
FOX
GATA
GATA
NKX
ARID
YY
HES
FOX
IPF
HNF
POU1F1
BRCA
CEBP
GATA
GR
PEBP
HNF
FOX
ZNF354C
GATA
GR
IPF
FOX
TBP
CEBP
RFX
PR
THAP
FOS
TBP
EBP
TBP
SRF
GATA
YY
HNF
SOX
CBF
JUN
TFIID
AP
TFIID
AP
RXR
HNF
HOX
TBP
AP
HNF
NF
CDX
FREACT
EN
GATA
ETF
ARI
PAX
RUNX
POU
MEF
ER
HLTF
HLTF
LF
YY
SRF
SP
CDX

Appendix A3

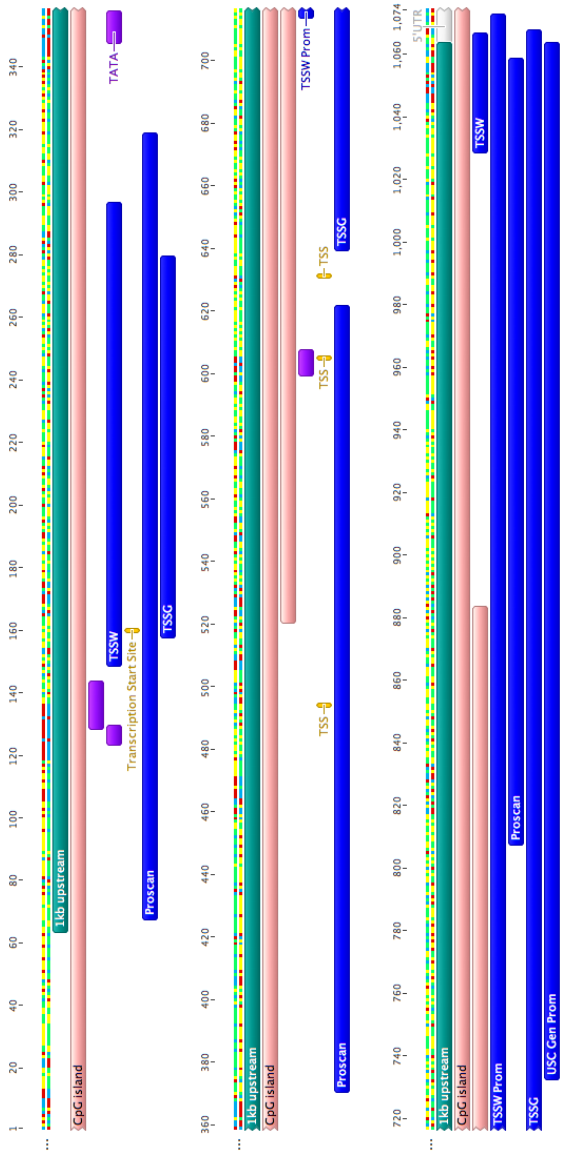


Figure A3.1: The human *ZIC2* putative promoter. A promoter region is predicted to occur within 1 kb upstream of the human *ZIC2* 5'UTR. The region is predicted to contain multiple CpG islands (salmon pink), TATA boxes (purple) and transcription start sites (yellow), all hallmarks of active promoters (Akan and Deloukas, 2008; Antequera and Bird, 1999). Blue: promoter regions predicted by Proscan, TSSG, TSSW and the UCSC Genome Browser (Prestridge, 1995; Rosenbloom *et al.*, 2015; Solovyev, 2008; Solovyev and Salamov, 1997; Solovyev and Shahmuradov, 2003).

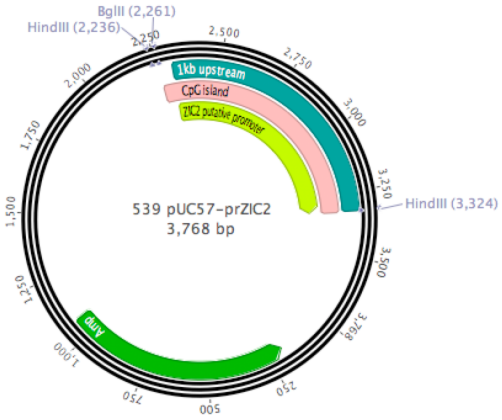
a)

*Bgl*II

AGATCTTTAAAGGGGCGGTTTGACCGGGGGGGCCCGGCCTCGAGCTGGAGGGAGGGAGGGAGG
CCGGGGCGGGAGACTAGGGGGTGCGGGGGGAGGGGAGAGGAAAAGGAGGAGACAAAAAATAAA
AAATAAAAGGCTGCCGCTGCAGCGTTGGCGGCGCCCATCGAAATCAACGGAGGCGGTGGCGAAC
GCAGCCCACCGCAGCCGAGACCTGGGAGCCCGCCTGGGCCTCACACTCCCTCGGGTCGCGGAC
TGCGCTGGGTCCACGCGGCGCGGTCACTAGTTCCGGGCCCAGCGCCCAGGCCCGACCGGCGG
GAGGGAGGAGGCGAGCGAGAGATCACTTTTTTATTGTTGTTATTGTTTTTCAGCCCAGCCTCCCCC
TCCTCCCCTCCCCCTGCTCGCTTTCTCCCCTCCCACATCCCCCTCCCCCCTACTCCCCCGCCTCC
TCCTCCGGCACAACTTAAAGAAAGGGGGAGCGGCGCGGCTGCTGCCTTCATCTGGGGAAATTCG
TGGCCACTGCAAGTTTACTACGCGAGGCGCAGCCAATGCCAAGCGCCGAGGCCGAGGAGGGCTA
AACACTGCGGCCGCGGCTCCGAACAATAACCGCCGCGCGCGGGGCGGCGCGAGTAGGGCCGCG
GGGGAGGGAGCTGTCGCCGCAGAGCGCCGCGGAGAGGACGCCTGGACTCCGCCTGCCGCCCC
GGCGCCCCCGCCGCGGTCAGGTGGAGCCGCTGGTGCGCTGGCCCCGGGTGCCGAGCGCGGAG
CTCGCCTCGGTCCTCTCCTCGCCCGTCTGCCTGGCGTGCGCACGGCCGCCGCGGTTGTGACTG
CATTCTACCGGCGCTGCTCGGTGCGGCCGGGCTCCGGGTCCGCTGGGCGTGCGAGTGAGTGTG
TGCGTGCGCGCGCGGGTGCGCGCAGGGGTGGGGGCTGCGGCGCGGCGCTCGCCCCCCGGTAG
CCCCCCTCTCCTCTCGGCTCCCCACACCTCCCTCCGGCTCCCTCCTCCCGCCCGCCCTCCCCTG
CCCTGCCCGCCCGCCCCCGCCGCAGCTCCTTTAATACACTTTGGTTCTCCGAAGCTT

*Hind*III

b)



c)

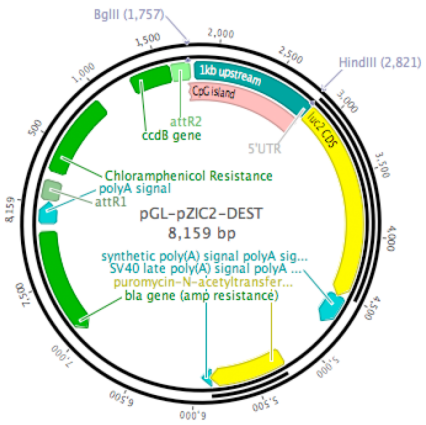


Figure A3.2: Attempts to clone the putative *ZIC2* promoter. Previous attempts by myself to amplify the putative *ZIC2* promoter for cloning were unsuccessful. As such, **(a)** the putative *ZIC2* promoter sequence with *Bgl*II and *Hind*III restriction sites flanking either end **(b)** was synthesised into pUC57 by Genscript. The resulting plasmid and pGL-DEST were digested with *Bgl*II and *Hind*III to release the *ZIC2* promoter and *ccdb* resistance fragments. Ligation of the digested *ZIC2* promoter fragment into the digested pGL-DEST backbone was attempted to create **(c)** pGl-pZIC2-DEST, however no successful constructs were able to be isolated.

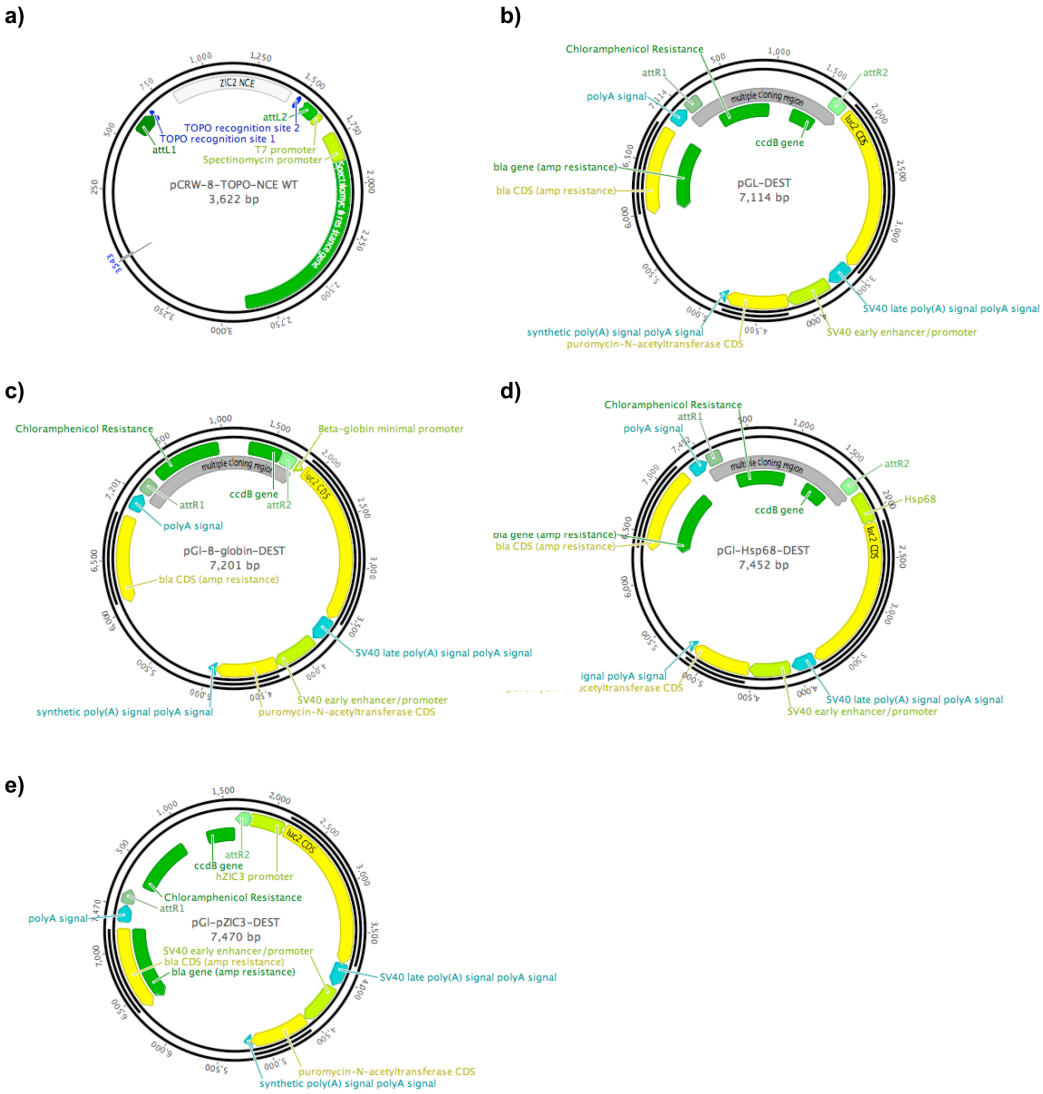


Figure A3.3: Plasmid maps. (a) pCRW-8-TOPO-NCE, **(b)** pGL-DEST with no promoter, **(c)** pGL-β-globin-DEST: pGL-DEST with a β-globin promoter cloned between attR2 and *luc* via *Bgl*II and *Hind*III sites, **(d)** pGL-Hsp68-DEST: pGL-DEST with a Hsp68 promoter cloned between attR2 and *luc* via *Bgl*II and *Hind*III sites, **(e)** pGL-pZIC3-DEST: pGL-DEST with a ZIC3 promoter cloned between attR2 and *luc* via *Bgl*II and *Hind*III sites.

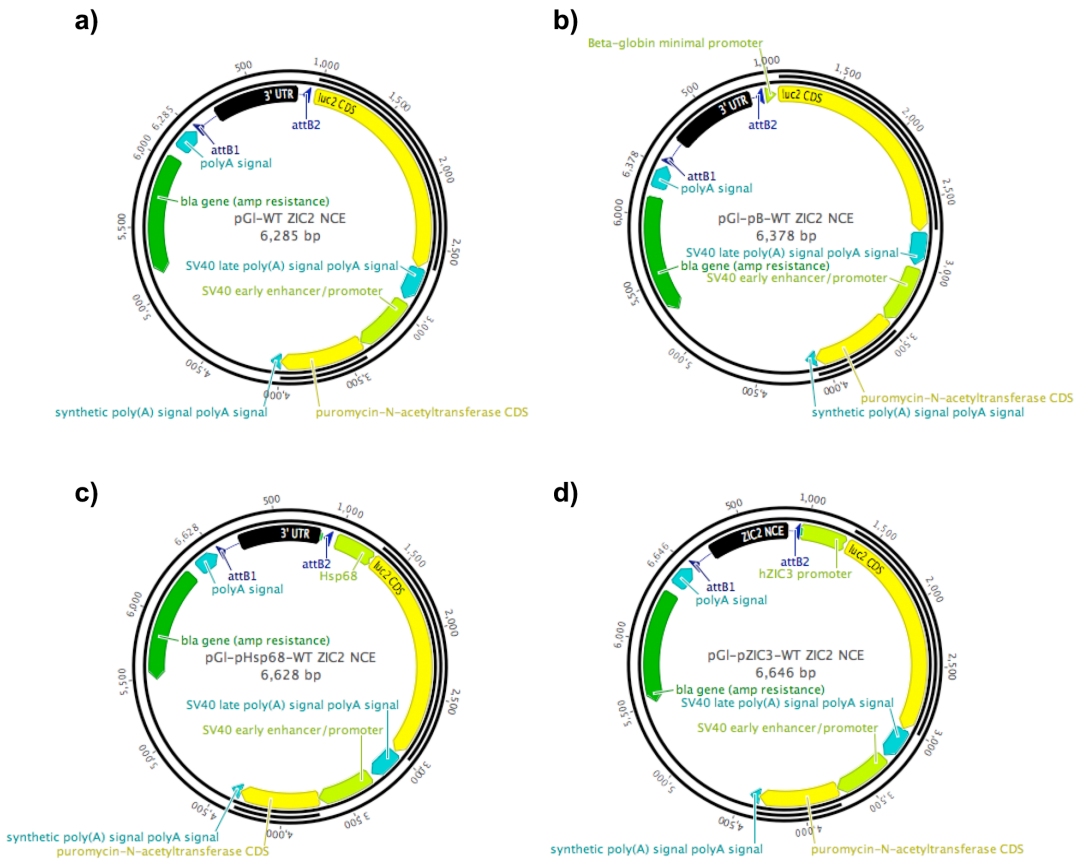


Figure A3.4: Plasmid maps. (a) pGl-WT *ZIC2* NCE with no promoter, **(b)** pGl-β-*globin*-WT *ZIC2* NCE with a *β-globin* promoter, **(c)** pGl-Hsp68-WT *ZIC2* NCE with a *Hsp68* promoter, **(d)** pGl-pZIC3-WT *ZIC2* NCE with a ZIC3 promoter. The six HPE-associated SNVs were introduced into the ZIC2 NCE in the pGl-pZIC3-WT *ZIC2* NCE constructs to create pGl-pZIC3-M6 *ZIC2* NCE (M1-M6).

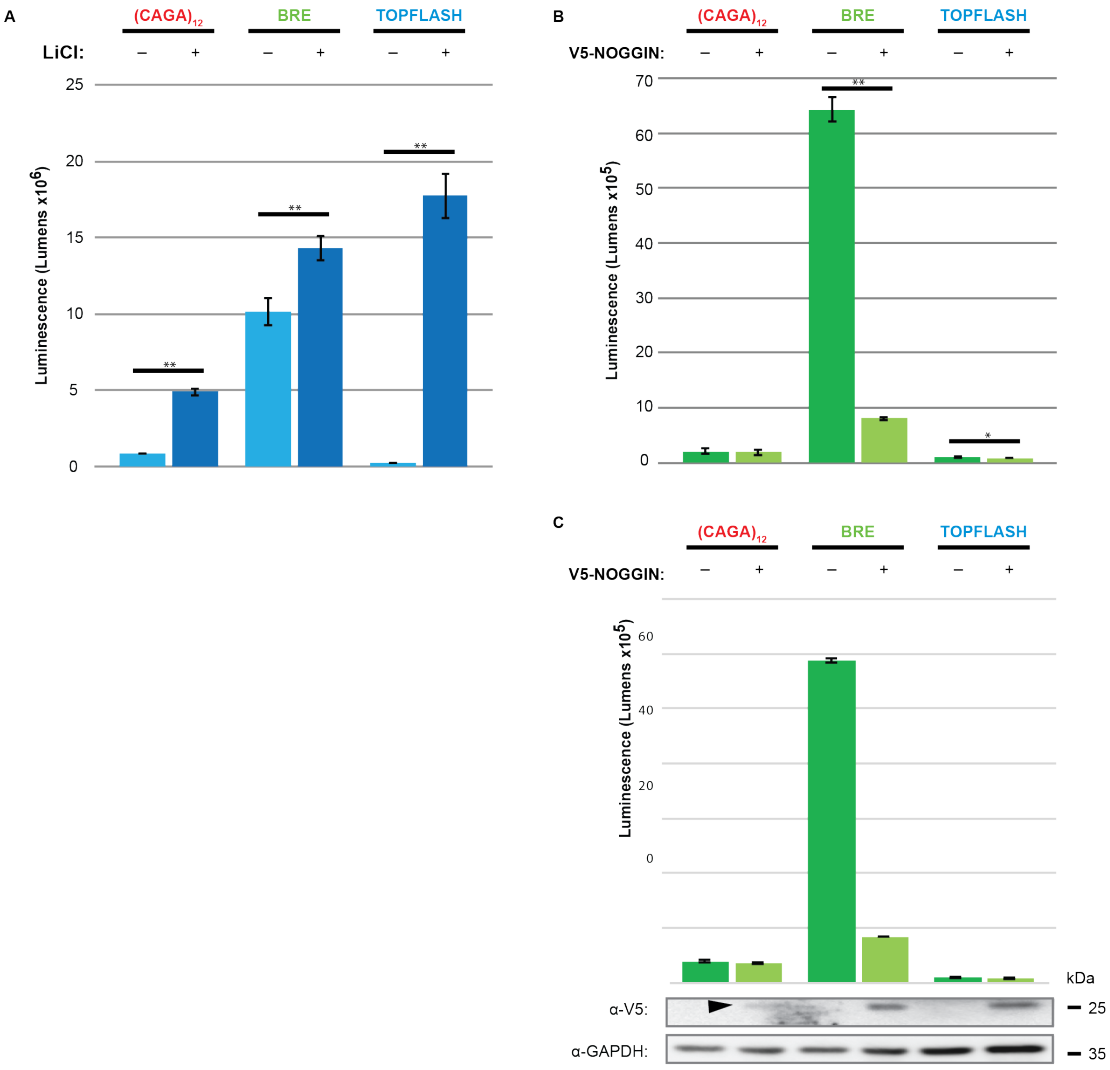


Figure A3.5: The addition of LiCl and V5-NOGGIN alters the signalling environment of HEK293T cells. (a) The addition of LiCl stimulates WNT signalling in HEK293T cells and also decreases BMP and NODAL signalling to create a medium NODAL, high BMP and high WNT environment. **(b)** The addition of V5-NOGGIN inhibits BMP signalling in HEK293T cells and also decreases WNT signalling. It has no effect on NODAL signalling, to create a low NODAL, mediium BMP and low WNT environment. HEK293T cells were transfected with the NODAL reporter $(CAGA)_{12}$, BMP reporter BRE or WNT reporter TOPFLASH, as well as V5-NOGGIN to inhibit BMP signalling, or an empty control vector (V5-DEST). Cells were cultured in DMEM with LiCl to promote WNT signalling, or DMEM on its own as a control. Reporter activity was measured as luminescence. Error bars = S.E.M of three external repeats; **: p<0.01, Student's T-Test. **(c)** Representative BMP signalling experiment and western blots. A single representative experiment is shown, with error bars = S.D., N=3 internal repeats. Western blots correspond to the transfections shown. α-GAPDH blots are included as loading controls

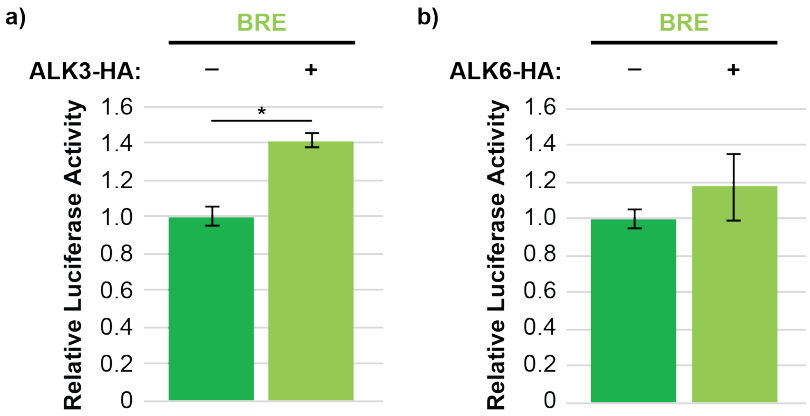


Figure A3.6: BMP signalling is already high in HEK293T cells. Constitutively active forms of the BMP receptors ALK3-HA and ALK6-HA were gifted from the Miyanzono laboratory, University of Tokyo. HEK293T cells were transfected with the BMP reporter BRE as well as **(a)** ALK3-HA or **(b)** ALK6-HA to stimulate BMP signalling, or an empty control vector (V5-DEST). Reporter activity was measured as luminescence and fold values calculated relative to unstimulated cells. Whilst ALK3-HA did generate a significant fold increase in BMP signalling, the increase seen was minor compared to that seen in Nodal signalling when stimulated with ALK4-HA, and WNT signalling when stimulated with LiCl. It can therefore be concluded that BMP signalling is already at its peak in HEK293T cells. Error bars = S.E.M of three external repeats; *: $p<0.05$, Student's T-Test.

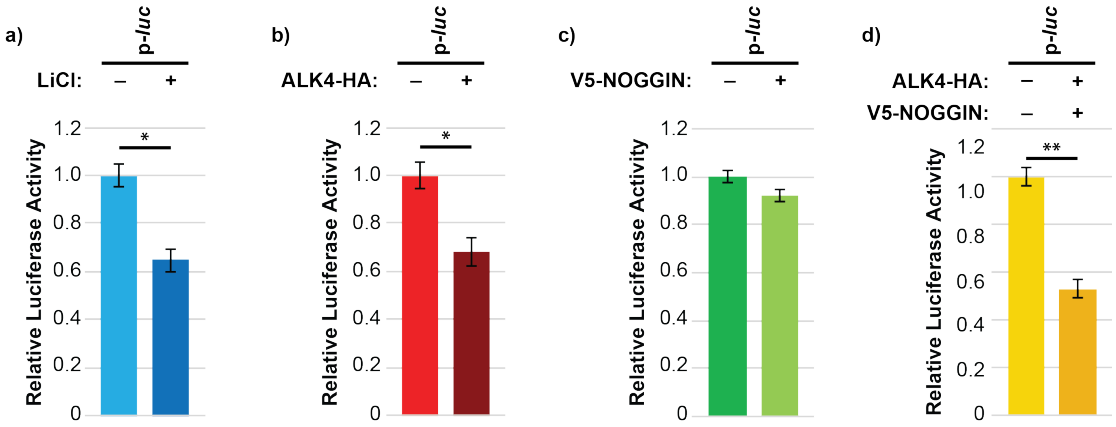


Figure A3.7: The activity of the *ZIC3* promoter (pZIC3) changes in different signalling environments. HEK293T cells were transfected with the empty reporter construct pGl-pZIC3-DEST (p-*luc*) and: incubated in **(a)** LiCl to stimulate WNT signalling, **(b)** transfected with ALK4-HA to stimulate NODAL signalling or an empty control vector (V5-DEST), **(c)** V5-NOGGIN to inhibit BMP signalling or V5-DEST, or **(d)** both ALK4-HA and V5-NOGGIN to stimulate a high NODAL/low BMP environment or V5-DEST. Reporter activity was measured as luminescence and fold values calculated relative to p-*luc* without signalling manipulation. Error bars = S.E.M of three external repeats; *: $p<0.05$, **: $p>0.01$, Student's T-test.

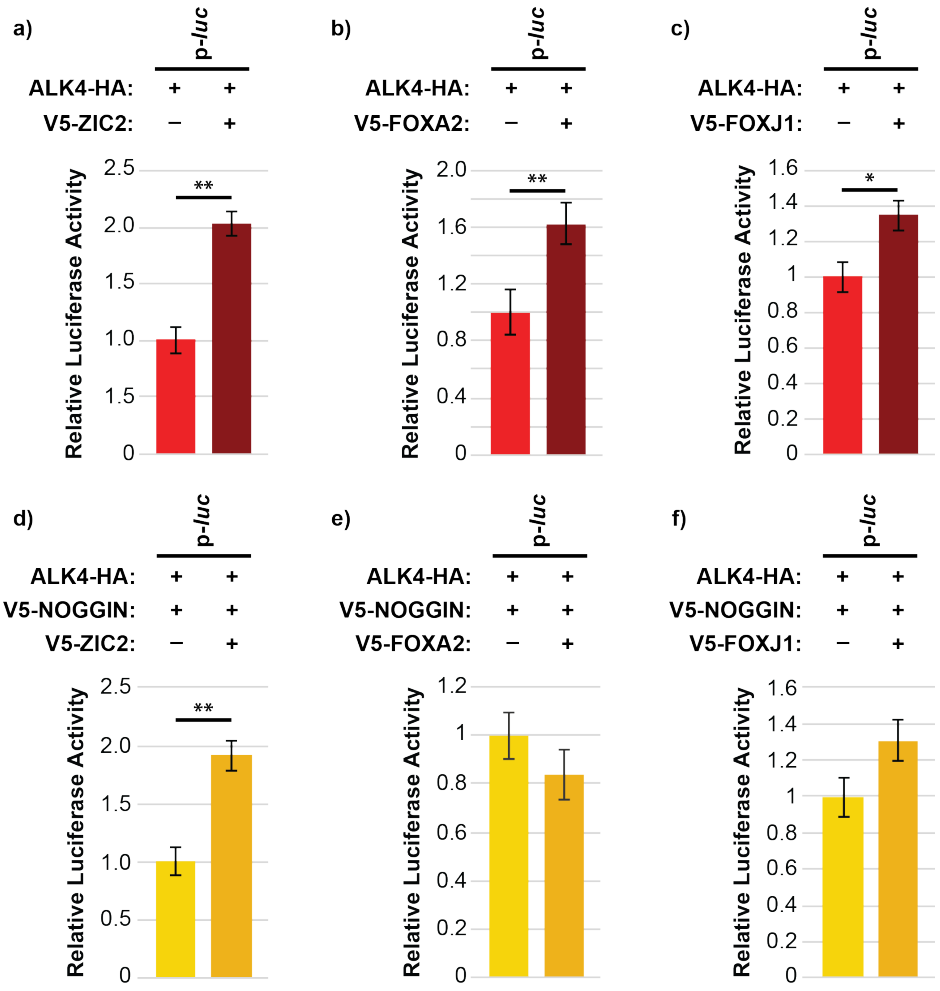
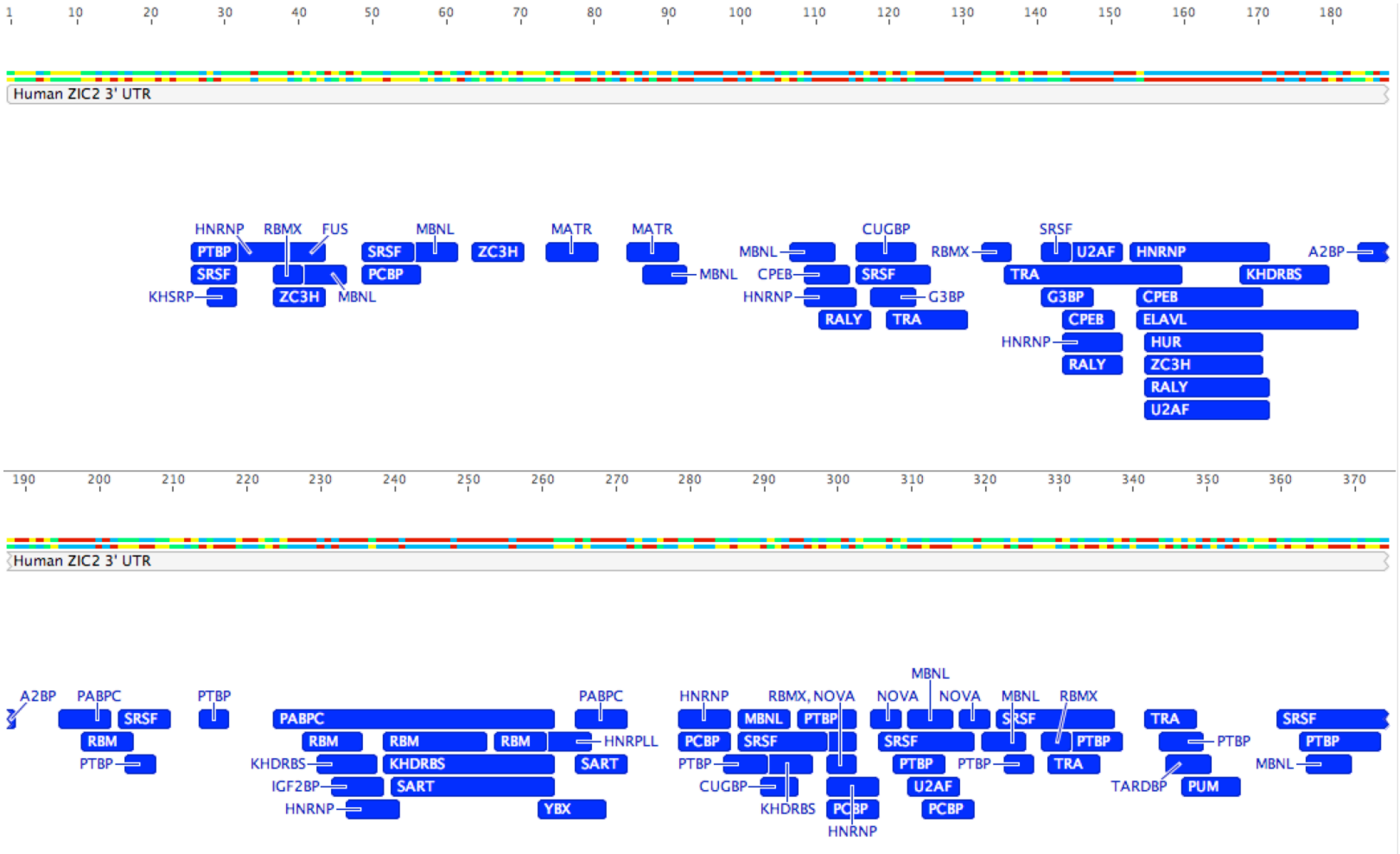
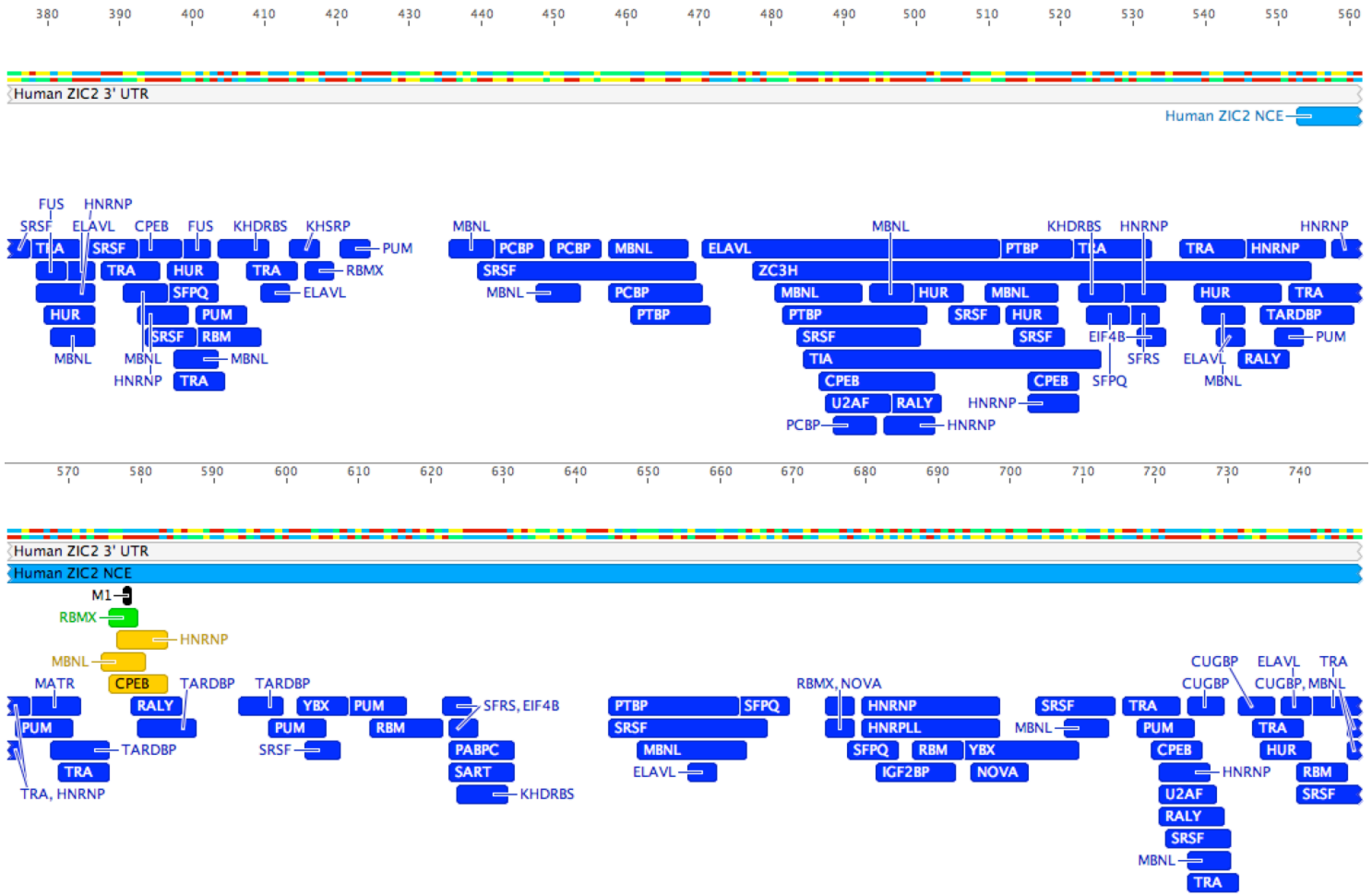


Figure A3.8: The addition of ZIC2, FOXA2 and FOXJ1 alter *ZIC3* promoter activity (pZIC3) in different signalling environments. HEK293T cells were transfected with the empty reporter construct pGl-pZIC3-DEST (p-*luc*) and **(a-c)** transfected with ALK4-HA to stimulate NODAL signalling or **(d-f)** both ALK4-HA and V5-NOGGIN to stimulate a high NODAL/low BMP environment. Cells were also transfected with either **(a, d)** V5-ZIC2, **(b, e)** V5-FOXA2, **(c, f)** V5-FOXJ1 or an empty control vector (V5-DEST). Reporter activity was measured as luminescence and fold values calculated relative to p-*luc* without signalling manipulation. Error bars = S.E.M of three external repeats; *: $p<0.05$, **: $p>0.01$, Student's T-test.

Appendix A4

Figure A4.1: Location of RBP binding sites predicted to occur in the *ZIC2* 3'UTR. Gold, mutations that disappear when the corresponding SNVs are introduced; green, binding sites created when the corresponding SNVs are introduced; and blue, binding sites that remain unchanged in both the wildtype and mutated 3'UTR. Binding sites were predicted by RBPmap database (Paz et al. 2014) and the RBPDB database (Berglund et al. 2008). HPE-associated SNVs are annotated in black.


1 10 20 30 40 50 60 70 80 90 100 110 120 130 140 150 160 170 180
Human ZIC2 3' UTR
HNRNP RBMX FUS MBNL MATR MATR CUGBP SRSF
PTBP SRSF ZC3H SRSF PCBP ZC3H MATR MBNL MBNL CUGBP RBMX SRSF U2AF HNRNP A2BP
SRSF PCBP MBNL CPEB SRSF TRA KHDRBS
KHSRP ZC3H MBNL HNRNP G3BP G3BP CPEB
RALY TRA CPEB ELAVL
HNRNP HUR
RALY ZC3H
RALY
U2AF
190 200 210 220 230 240 250 260 270 280 290 300 310 320 330 340 350 360 370
Human ZIC2 3' UTR
A2BP PABPC PTBP PABPC HNRNP RBMX, NOVA NOVA MBNL NOVA MBNL RBMX
SRSF PABPC MBNL PTBP SRSF SRSF TRA SRSF
RBM RBM RBM HNRPLL PCBP SRSF PTBP PTBP PTBP
PTBP KHDRBS KHDRBS SART PTBP PTBP TRA PTBP PTBP
IGF2BP SART CUGBP U2AF MBNL
HNRNP YBX KHDRBS PCBP PCBP TARDBP PUM
HNRNP


380 390 400 410 420 430 440 450 460 470 480 490 500 510 520 530 540 550 560
Human ZIC2 3' UTR
Human ZIC2 NCE
FUS HNRNP SRSF ELAVL CPEB FUS KHDRBS KHSRP PUM RBMX ELAVL
MBNL PCBP PCBP MBNL ELAVL PTBP TIA TRA HNRNP HNRNP SRSF ZC3H
MBNL KHDRBS HNRNP HNRNP HUR TRA TARDBP PUM EIF4B SFRS ELAVL RALY MBNL
TRA HUR SFPQ PUM SRSF RBM MBNL HNRNP TRA PTBP SRSF HUR SRSF TIA CPEB CPEB SFPQ U2AF RALY HNRNP PCBP HNRNP
570 580 590 600 610 620 630 640 650 660 670 680 690 700 710 720 730 740
Human ZIC2 3' UTR
Human ZIC2 NCE
M1 RBMX HNRNP MBNL MATR CPEB TARDBP TARDBP SFRS, EIF4B
RALY YBX PUM PUM RBM PABPC SART KHDRBS
PUM TARDBP SRSF TRA TRA, HNRNP
PTBP SFPQ SRSF MBNL ELAVL
RBMX, NOVA HNRNP HNRPLL SFPQ RBM YBX IGF2BP NOVA
SRSF MBNL
CUGBP ELAVL TRA CUGBP CUGBP, MBNL
TRA PUM CPEB HNRNP U2AF RALY SRSF MBNL TRA
TRA HUR RBM SRSF

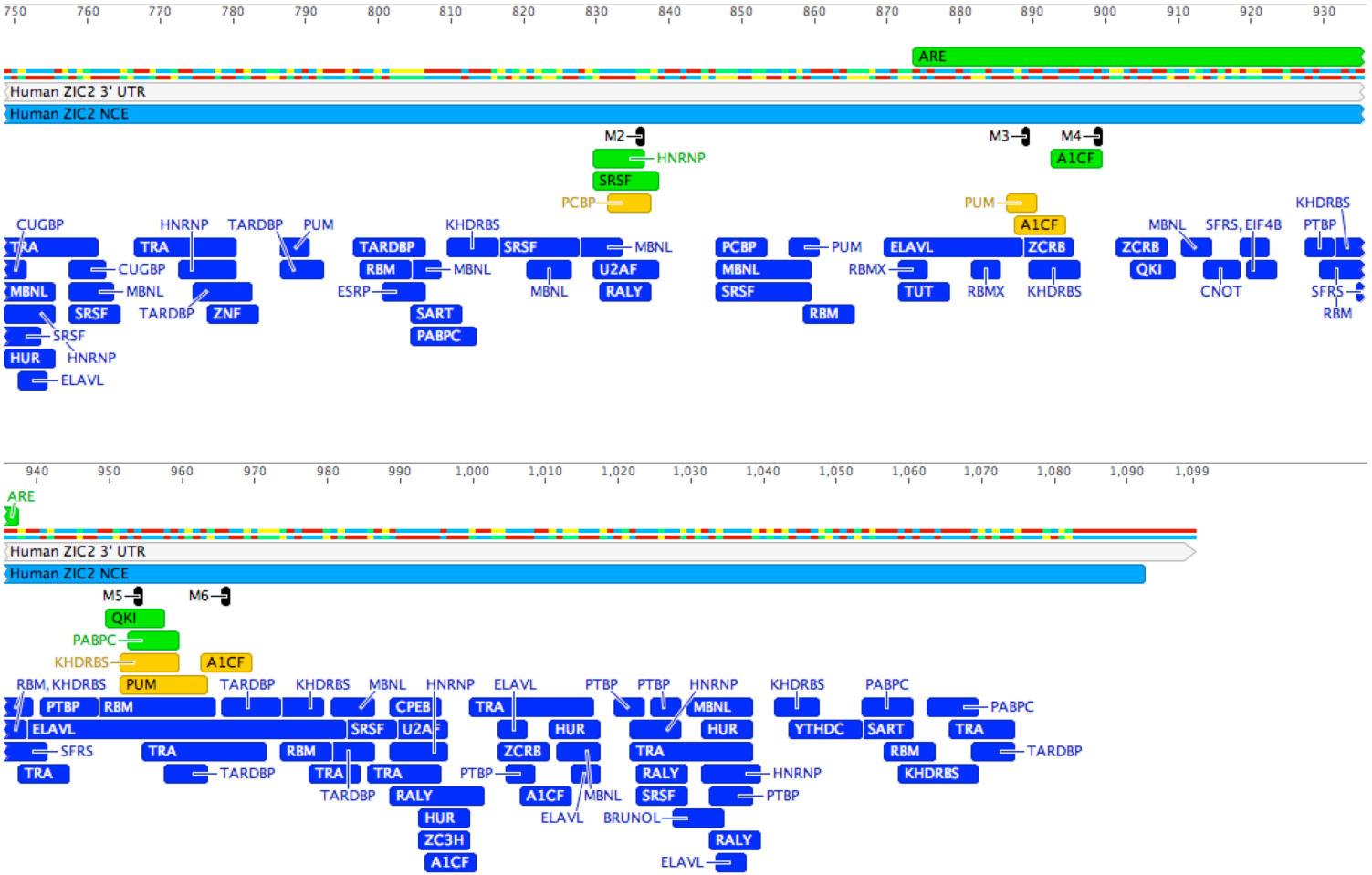
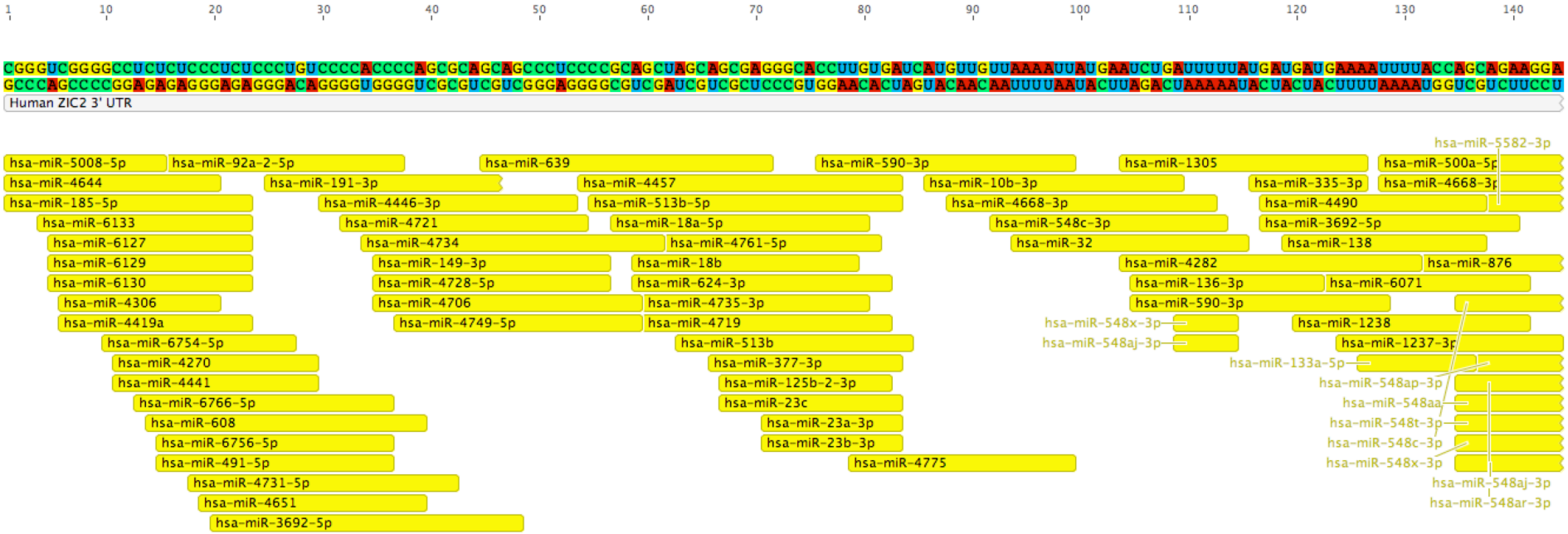
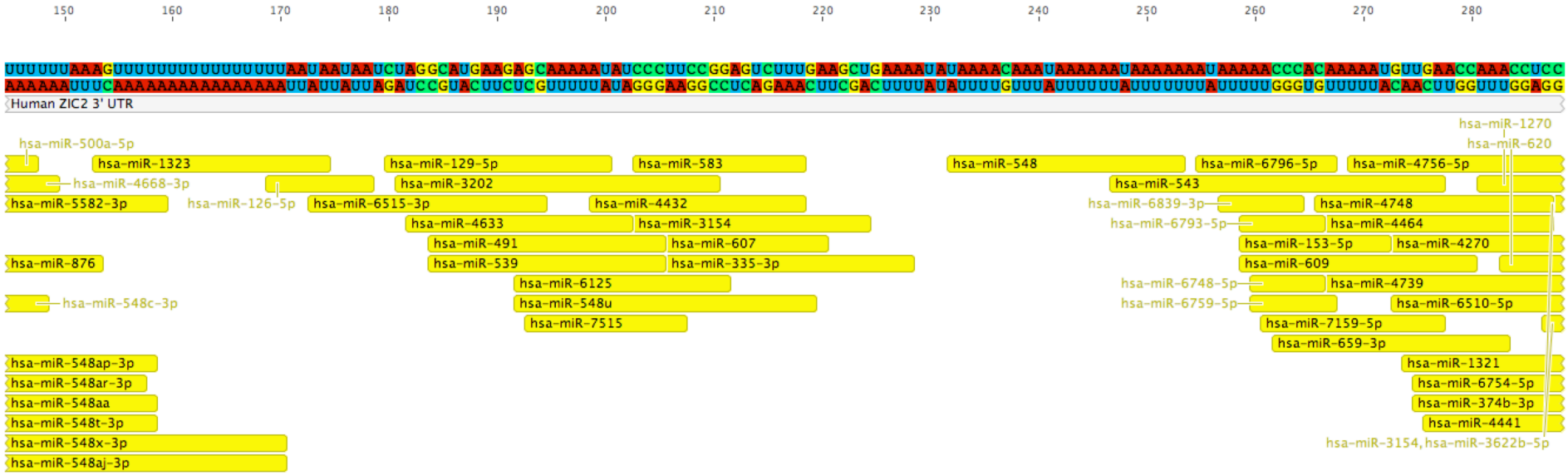
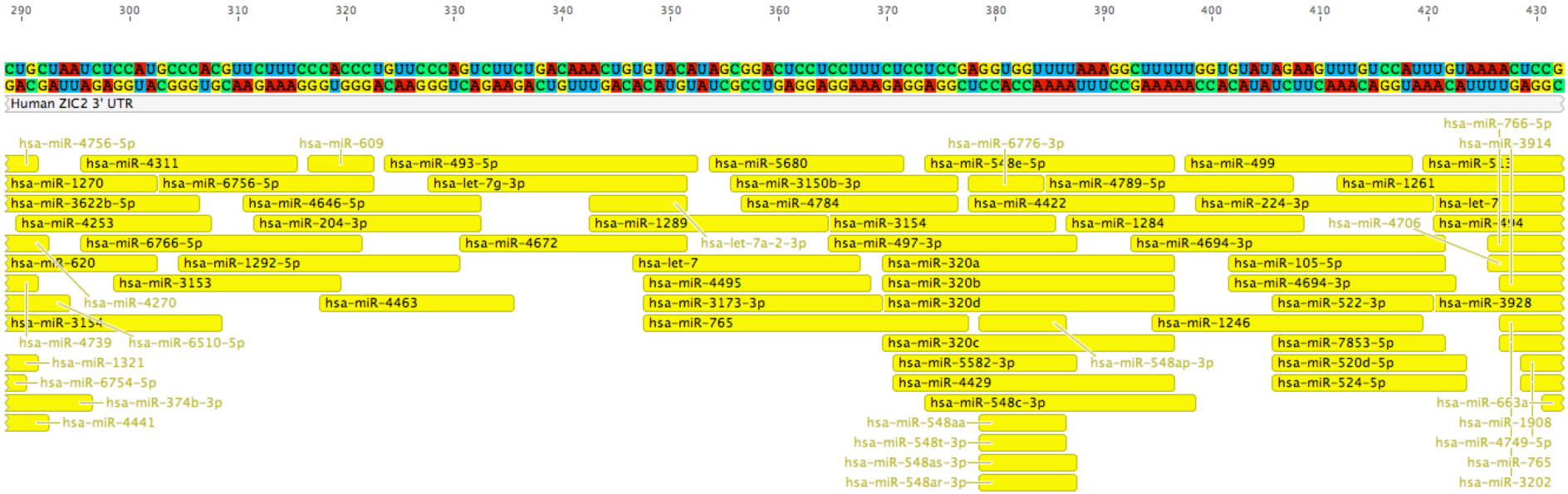
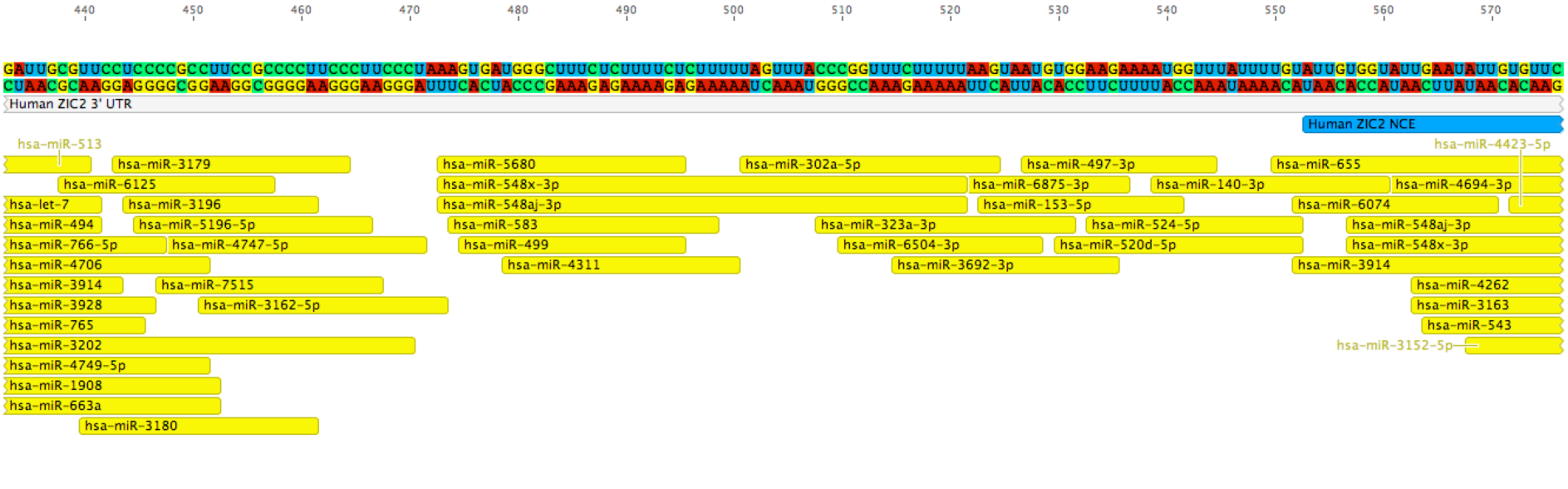

750 760 770 780 790 800 810 820 830 840 850 860 870 880 890 900 910 920 930
ARE
Human ZIC2 3' UTR
Human ZIC2 NCE
M2 M3 M4
HNRNP
A1CF
SRSF
PCBP
PUM
KHDRBS
CUGBP
HNRNP
TARDBP
PUM
KHDRBS
MBNL
SFRS, EIF4B
PTBP
TRA
TRA
TARDBP
TARDBP
SRSF
MBNL
PCBP
PUM
ELAVL
ZCRB
ZCRB
CUGBP
RBM
MBNL
MBNL
U2AF
MBNL
RBMX
QKI
MBNL
SRSF
MBNL
TARDBP
ZNF
ESRP
RALY
SRSF
TUT
RBMX
KHDRBS
CNOT
SFRS
SART
RBM
SRSF
PABPC
RBM
HUR
HNRNP
ELAVL
940 950 960 970 980 990 1,000 1,010 1,020 1,030 1,040 1,050 1,060 1,070 1,080 1,090 1,099
ARE
Human ZIC2 3' UTR
Human ZIC2 NCE
M5 M6
QKI
PABPC
KHDRBS
A1CF
RBM, KHDRBS
PUM
TARDBP
KHDRBS
MBNL
HNRNP
ELAVL
PTBP
PTBP
HNRNP
KHDRBS
PABPC
PTBP
RBM
CPEB
TRA
MBNL
PABPC
ELAVL
SRSF
U2A
HUR
HUR
YTHDC
SART
TRA
SFRS
TRA
RBM
ZCRB
TRA
RBM
TARDBP
TRA
TARDBP
TRA
TRA
PTBP
RALY
HNRNP
KHDRBS
TARDBP
RALY
A1CF
MBNL
SRSF
PTBP
ELAVL
BRUNOL
HUR
ZC3H
A1CF
RALY
ELAVL

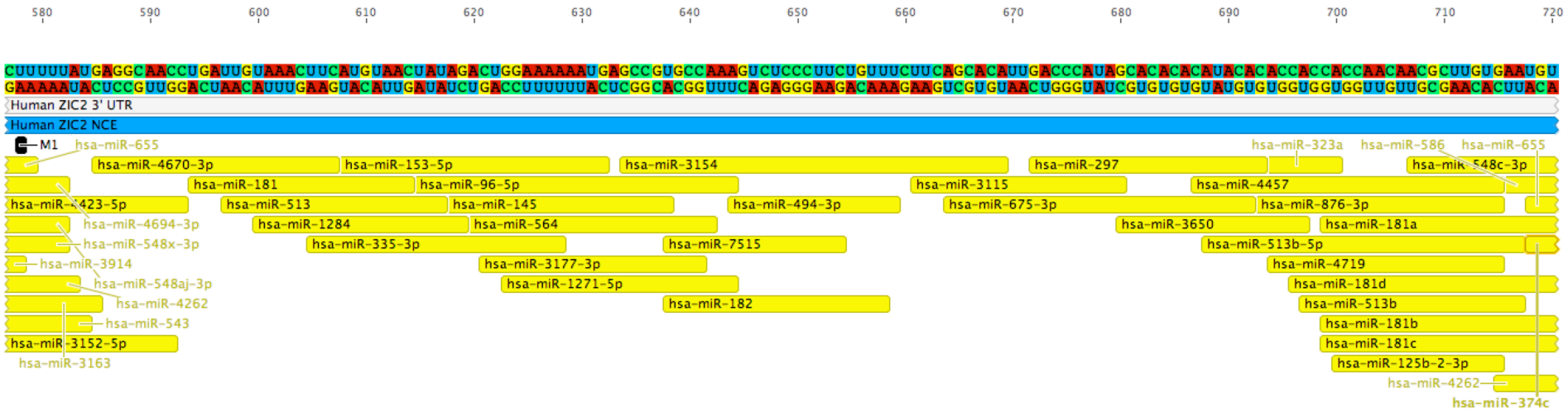
Figure A4.2: Location of miRNA binding sites predicted to occur in the *ZIC2* 3'UTR. Binding sites (yellow) were predicted by miRanda (John et al. 2004; Betel et al. 2008), miRNASNP (Gong et al. 2015) and DIANA-microT-CDS (Reczko et al. 2012; Paraskevopoulou et al. 2013). HPE-associated SNVs are annotated in black.

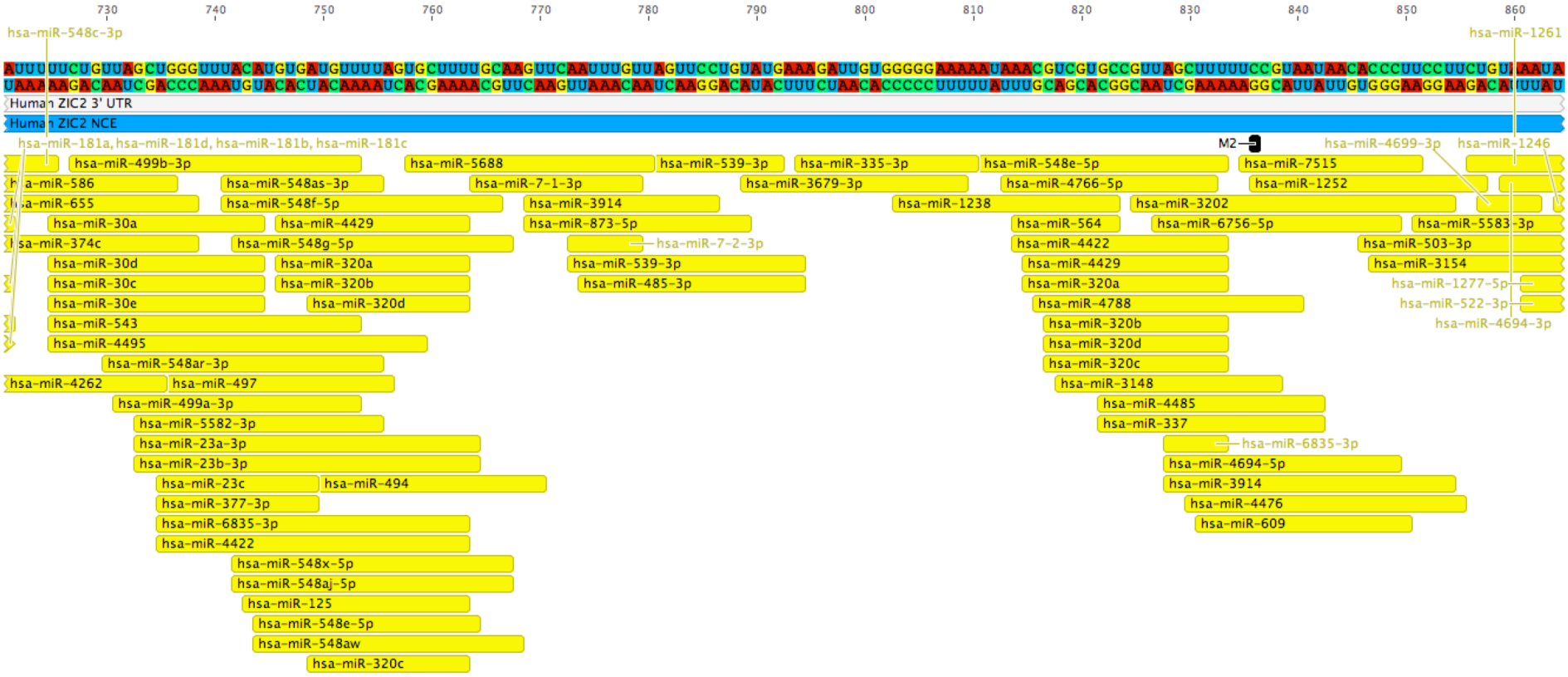

1
10
20
30
40
50
60
70
80
90
100
110
120
130
140
CGGGUCGGGGCCUCUCUCCCUCUCCCUGUCCCCACCCCAGCGCAGCAGCCCCUCCCCGCAGCUAGCAGCGAGGGCACCUUGUGAUCAUGUUGUUAAAAUUAUGAAUCUGAUUUUUAUGAUGAUGAAAAUUUUACCAGCAGAAGGA
GCCCAGCCCCGGAGAGAGGGAGAGGGACAGGGGUGGGGUCGCGUCGUCGGGAGGGGCGUCGAUCGUCGCUCCCGUGGAACACUAGUACAACAAUUUUAAUACUUAGACUAAAAAUACUACUACUUUUAAAAUGGUCGUCUUCCU
Human ZIC2 3' UTR
hsa-miR-5582-3p
hsa-miR-5008-5p
hsa-miR-92a-2-5p
hsa-miR-639
hsa-miR-590-3p
hsa-miR-1305
hsa-miR-500a-5p
hsa-miR-4644
hsa-miR-191-3p
hsa-miR-4457
hsa-miR-10b-3p
hsa-miR-335-3p
hsa-miR-4668-3p
hsa-miR-185-5p
hsa-miR-4446-3p
hsa-miR-513b-5p
hsa-miR-4668-3p
hsa-miR-4490
hsa-miR-6133
hsa-miR-4721
hsa-miR-18a-5p
hsa-miR-548c-3p
hsa-miR-3692-5p
hsa-miR-6127
hsa-miR-4734
hsa-miR-4761-5p
hsa-miR-32
hsa-miR-138
hsa-miR-6129
hsa-miR-149-3p
hsa-miR-18b
hsa-miR-4282
hsa-miR-876
hsa-miR-6130
hsa-miR-4728-5p
hsa-miR-624-3p
hsa-miR-136-3p
hsa-miR-6071
hsa-miR-4306
hsa-miR-4706
hsa-miR-4735-3p
hsa-miR-590-3p
hsa-miR-4419a
hsa-miR-4749-5p
hsa-miR-4719
hsa-miR-548x-3p
hsa-miR-1238
hsa-miR-6754-5p
hsa-miR-513b
hsa-miR-548aj-3p
hsa-miR-1237-3p
hsa-miR-4270
hsa-miR-377-3p
hsa-miR-133a-5p
hsa-miR-4441
hsa-miR-125b-2-3p
hsa-miR-548ap-3p
hsa-miR-6766-5p
hsa-miR-23c
hsa-miR-548aa
hsa-miR-608
hsa-miR-23a-3p
hsa-miR-548t-3p
hsa-miR-6756-5p
hsa-miR-23b-3p
hsa-miR-548c-3p
hsa-miR-491-5p
hsa-miR-4775
hsa-miR-548x-3p
hsa-miR-4731-5p
hsa-miR-548aj-3p
hsa-miR-4651
hsa-miR-548ar-3p
hsa-miR-3692-5p

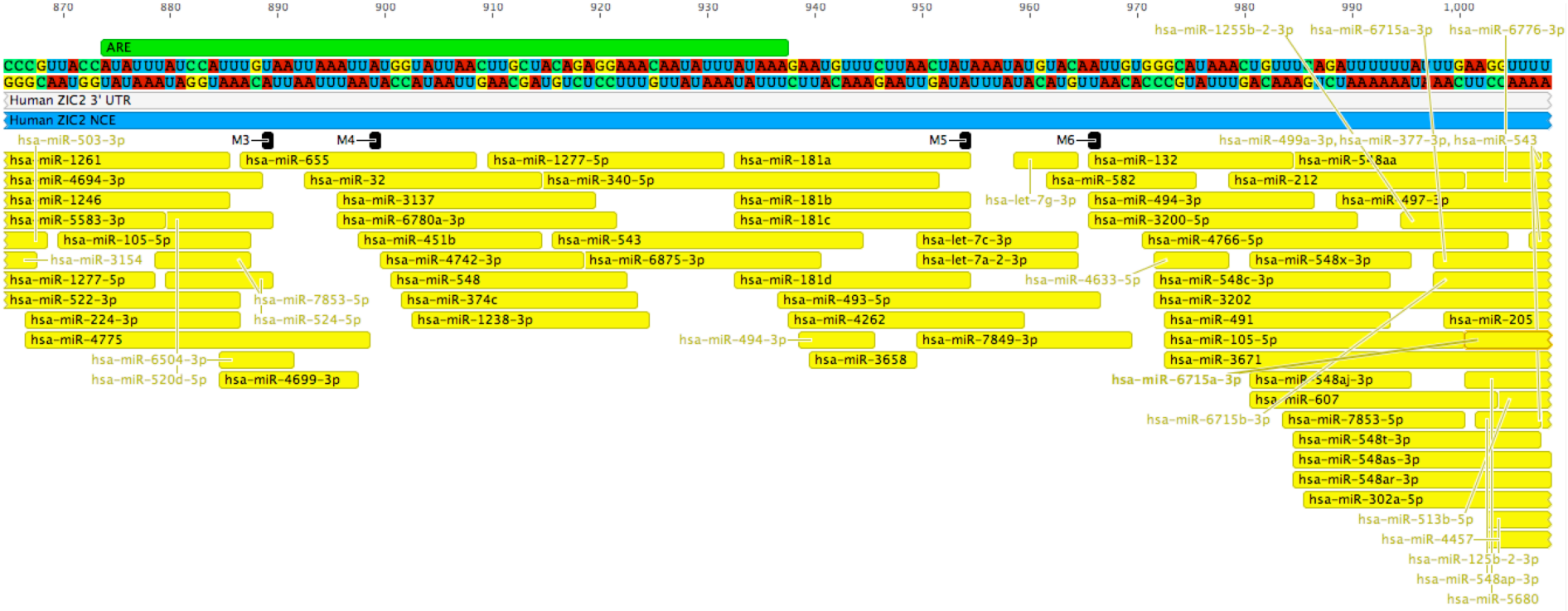

150
160
170
180
190
200
210
220
230
240
250
260
270
280
Human ZIC2 3' UTR
hsa-miR-1270
hsa-miR-620
hsa-miR-500a-5p
hsa-miR-1323
hsa-miR-129-5p
hsa-miR-583
hsa-miR-548
hsa-miR-6796-5p
hsa-miR-4756-5p
hsa-miR-4668-3p
hsa-miR-3202
hsa-miR-543
hsa-miR-5582-3p
hsa-miR-126-5p
hsa-miR-6515-3p
hsa-miR-4432
hsa-miR-6839-3p
hsa-miR-4748
hsa-miR-4633
hsa-miR-3154
hsa-miR-6793-5p
hsa-miR-4464
hsa-miR-491
hsa-miR-607
hsa-miR-153-5p
hsa-miR-4270
hsa-miR-876
hsa-miR-539
hsa-miR-335-3p
hsa-miR-609
hsa-miR-6125
hsa-miR-6748-5p
hsa-miR-4739
hsa-miR-548c-3p
hsa-miR-548u
hsa-miR-6759-5p
hsa-miR-6510-5p
hsa-miR-7515
hsa-miR-7159-5p
hsa-miR-659-3p
hsa-miR-548ap-3p
hsa-miR-1321
hsa-miR-548ar-3p
hsa-miR-6754-5p
hsa-miR-548aa
hsa-miR-374b-3p
hsa-miR-548t-3p
hsa-miR-4441
hsa-miR-548x-3p
hsa-miR-3154,hsa-miR-3622b-5p
hsa-miR-548aj-3p

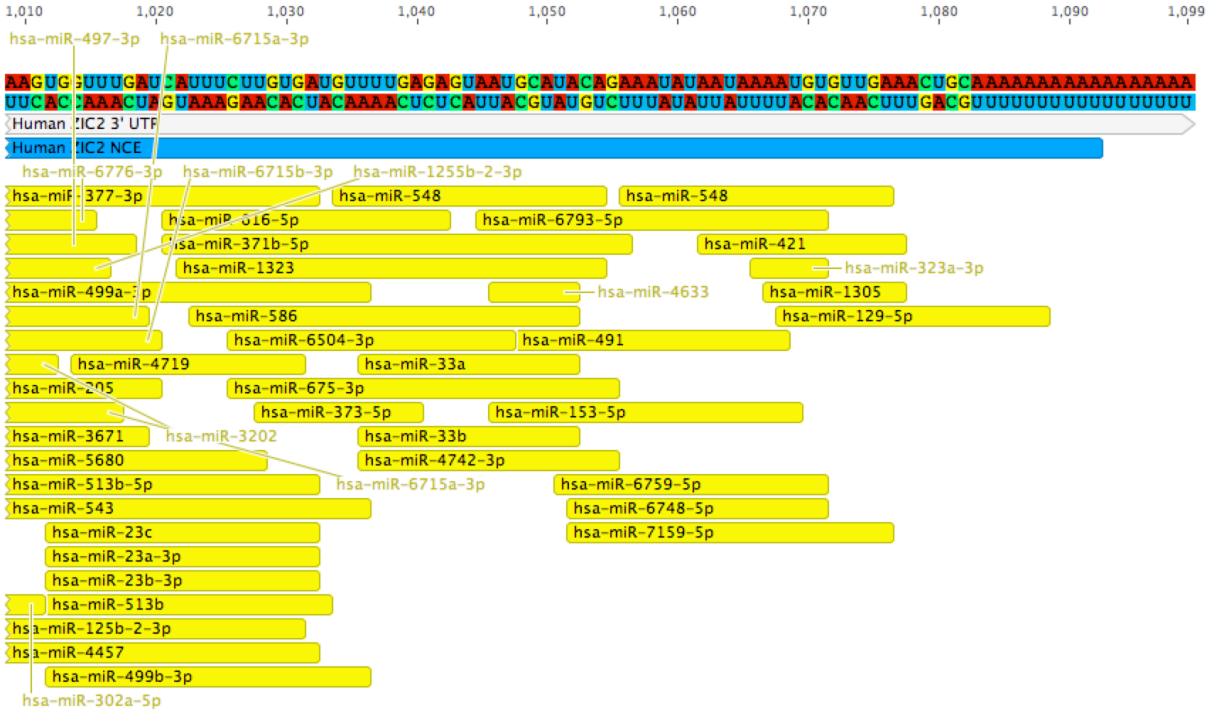

290
300
310
320
330
340
350
360
370
380
390
400
410
420
430
CUGCUAAUCUCCAUGCCCACGUUCUUUCCCACCCUGUUCCCAGUCUUCUGACAAACUGUGUACAUAGCGGACUCCUCCUUUCUCCUCCGAGGUGGUUUUAAAGGCUUUUUGGUGUAUAGAAGUUUGUCCAUUUGUAAAACUCCG
GACGAUUAGAGGUACGGGUGCAAGAAAGGGUGGGACAAGGGUCAGAAGACUGUUUGACACAUGUAUCGCCUGAGGAGGAAAGAGGAGGCUCCACCAAAAUUUCCGAAAAACCACAUAUCUUCAAACAGGUAAACAUUUUGAGGC
Human ZIC2 3' UTR
hsa-miR-4756-5p
hsa-miR-609
hsa-miR-6776-3p
hsa-miR-766-5p
hsa-miR-3914
hsa-miR-4311
hsa-miR-493-5p
hsa-miR-5680
hsa-miR-499
hsa-miR-1270
hsa-miR-6756-5p
hsa-let-7g-3p
hsa-miR-3150b-3p
hsa-miR-4789-5p
hsa-miR-1261
hsa-miR-3622b-5p
hsa-miR-4646-5p
hsa-miR-4784
hsa-miR-4422
hsa-miR-224-3p
hsa-let-7
hsa-miR-4253
hsa-miR-204-3p
hsa-miR-1289
hsa-miR-3154
hsa-miR-1284
hsa-miR-4706
hsa-miR-6766-5p
hsa-miR-4672
hsa-let-7a-2-3p
hsa-miR-497-3p
hsa-miR-4694-3p
hsa-miR-620
hsa-miR-1292-5p
hsa-let-7
hsa-miR-320a
hsa-miR-105-5p
hsa-miR-3153
hsa-miR-4495
hsa-miR-320b
hsa-miR-4694-3p
hsa-miR-4270
hsa-miR-4463
hsa-miR-3173-3p
hsa-miR-320d
hsa-miR-522-3p
hsa-miR-3928
hsa-miR-765
hsa-miR-1246
hsa-miR-4739
hsa-miR-6510-5p
hsa-miR-320c
hsa-miR-7853-5p
hsa-miR-1321
hsa-miR-5582-3p
hsa-miR-548ap-3p
hsa-miR-520d-5p
hsa-miR-6754-5p
hsa-miR-4429
hsa-miR-524-5p
hsa-miR-374b-3p
hsa-miR-548c-3p
hsa-miR-663a
hsa-miR-4441
hsa-miR-548aa
hsa-miR-1908
hsa-miR-548t-3p
hsa-miR-4749-5p
hsa-miR-548as-3p
hsa-miR-765
hsa-miR-548ar-3p
hsa-miR-3202


440
450
460
470
480
490
500
510
520
530
540
550
560
570
Human ZIC2 3' UTR
Human ZIC2 NCE
hsa-miR-513
hsa-miR-4423-5p
hsa-miR-3179
hsa-miR-5680
hsa-miR-302a-5p
hsa-miR-497-3p
hsa-miR-655
hsa-miR-6125
hsa-miR-548x-3p
hsa-miR-6875-3p
hsa-miR-140-3p
hsa-miR-4694-3p
hsa-let-7
hsa-miR-3196
hsa-miR-548aj-3p
hsa-miR-153-5p
hsa-miR-6074
hsa-miR-494
hsa-miR-5196-5p
hsa-miR-583
hsa-miR-323a-3p
hsa-miR-524-5p
hsa-miR-548aj-3p
hsa-miR-766-5p
hsa-miR-4747-5p
hsa-miR-499
hsa-miR-6504-3p
hsa-miR-520d-5p
hsa-miR-548x-3p
hsa-miR-4706
hsa-miR-4311
hsa-miR-3692-3p
hsa-miR-3914
hsa-miR-3914
hsa-miR-7515
hsa-miR-4262
hsa-miR-3928
hsa-miR-3162-5p
hsa-miR-3163
hsa-miR-765
hsa-miR-543
hsa-miR-3202
hsa-miR-3152-5p
hsa-miR-4749-5p
hsa-miR-1908
hsa-miR-663a
hsa-miR-3180


580
590
600
610
620
630
640
650
660
670
680
690
700
710
720
Human ZIC2 3' UTR
Human ZIC2 NCE
M1
hsa-miR-655
hsa-miR-4670-3p
hsa-miR-153-5p
hsa-miR-3154
hsa-miR-297
hsa-miR-323a
hsa-miR-586
hsa-miR-655
hsa-miR-548c-3p
hsa-miR-181
hsa-miR-96-5p
hsa-miR-3115
hsa-miR-4457
hsa-miR-423-5p
hsa-miR-513
hsa-miR-145
hsa-miR-494-3p
hsa-miR-675-3p
hsa-miR-876-3p
hsa-miR-4694-3p
hsa-miR-1284
hsa-miR-564
hsa-miR-3650
hsa-miR-181a
hsa-miR-548x-3p
hsa-miR-335-3p
hsa-miR-7515
hsa-miR-513b-5p
hsa-miR-3914
hsa-miR-3177-3p
hsa-miR-4719
hsa-miR-548aj-3p
hsa-miR-1271-5p
hsa-miR-181d
hsa-miR-4262
hsa-miR-182
hsa-miR-513b
hsa-miR-543
hsa-miR-181b
hsa-miR-3152-5p
hsa-miR-181c
hsa-miR-3163
hsa-miR-125b-2-3p
hsa-miR-4262
hsa-miR-374c


730
740
750
760
770
780
790
800
810
820
830
840
850
860
hsa-miR-548c-3p
hsa-miR-1261
Human ZIC2 3' UTR
Human ZIC2 NCE
hsa-miR-181a, hsa-miR-181d, hsa-miR-181b, hsa-miR-181c
M2
hsa-miR-4699-3p
hsa-miR-1246
hsa-miR-499b-3p
hsa-miR-5688
hsa-miR-539-3p
hsa-miR-335-3p
hsa-miR-548e-5p
hsa-miR-7515
hsa-miR-586
hsa-miR-548as-3p
hsa-miR-7-1-3p
hsa-miR-3679-3p
hsa-miR-4766-5p
hsa-miR-1252
hsa-miR-655
hsa-miR-548f-5p
hsa-miR-3914
hsa-miR-1238
hsa-miR-3202
hsa-miR-30a
hsa-miR-4429
hsa-miR-873-5p
hsa-miR-564
hsa-miR-6756-5p
hsa-miR-5583-3p
hsa-miR-374c
hsa-miR-548g-5p
hsa-miR-7-2-3p
hsa-miR-4422
hsa-miR-503-3p
hsa-miR-30d
hsa-miR-320a
hsa-miR-539-3p
hsa-miR-4429
hsa-miR-3154
hsa-miR-30c
hsa-miR-320b
hsa-miR-485-3p
hsa-miR-320a
hsa-miR-1277-5p
hsa-miR-30e
hsa-miR-320d
hsa-miR-4788
hsa-miR-522-3p
hsa-miR-543
hsa-miR-320b
hsa-miR-4694-3p
hsa-miR-4495
hsa-miR-320d
hsa-miR-548ar-3p
hsa-miR-320c
hsa-miR-4262
hsa-miR-497
hsa-miR-3148
hsa-miR-499a-3p
hsa-miR-4485
hsa-miR-5582-3p
hsa-miR-337
hsa-miR-23a-3p
hsa-miR-6835-3p
hsa-miR-23b-3p
hsa-miR-4694-5p
hsa-miR-23c
hsa-miR-494
hsa-miR-3914
hsa-miR-377-3p
hsa-miR-4476
hsa-miR-6835-3p
hsa-miR-609
hsa-miR-4422
hsa-miR-548x-5p
hsa-miR-548aj-5p
hsa-miR-125
hsa-miR-548e-5p
hsa-miR-548aw
hsa-miR-320c


870
880
890
900
910
920
930
940
950
960
970
980
990
1,000
ARE
Human ZIC2 3' UTR
Human ZIC2 NCE
M3
M4
M5
M6
hsa-miR-503-3p
hsa-miR-1261
hsa-miR-4694-3p
hsa-miR-1246
hsa-miR-5583-3p
hsa-miR-105-5p
hsa-miR-3154
hsa-miR-1277-5p
hsa-miR-522-3p
hsa-miR-224-3p
hsa-miR-4775
hsa-miR-6504-3p
hsa-miR-520d-5p
hsa-miR-4699-3p
hsa-miR-655
hsa-miR-32
hsa-miR-3137
hsa-miR-6780a-3p
hsa-miR-451b
hsa-miR-4742-3p
hsa-miR-548
hsa-miR-7853-5p
hsa-miR-374c
hsa-miR-524-5p
hsa-miR-1238-3p
hsa-miR-1277-5p
hsa-miR-340-5p
hsa-miR-543
hsa-miR-6875-3p
hsa-miR-181a
hsa-miR-181b
hsa-miR-181c
hsa-miR-181d
hsa-miR-493-5p
hsa-miR-4262
hsa-miR-494-3p
hsa-miR-3658
hsa-let-7c-3p
hsa-let-7a-2-3p
hsa-miR-7849-3p
hsa-let-7g-3p
hsa-miR-4633-5p
hsa-miR-1255b-2-3p
hsa-miR-6715a-3p
hsa-miR-6776-3p
hsa-miR-499a-3p, hsa-miR-377-3p, hsa-miR-543
hsa-miR-132
hsa-miR-582
hsa-miR-494-3p
hsa-miR-3200-5p
hsa-miR-4766-5p
hsa-miR-548x-3p
hsa-miR-548c-3p
hsa-miR-3202
hsa-miR-491
hsa-miR-105-5p
hsa-miR-3671
hsa-miR-6715a-3p
hsa-miR-607
hsa-miR-6715b-3p
hsa-miR-7853-5p
hsa-miR-548t-3p
hsa-miR-548as-3p
hsa-miR-548ar-3p
hsa-miR-302a-5p
hsa-miR-513b-5p
hsa-miR-4457
hsa-miR-125b-2-3p
hsa-miR-548ap-3p
hsa-miR-5680
hsa-miR-212
hsa-miR-205


1,010
1,020
1,030
1,040
1,050
1,060
1,070
1,080
1,090
1,099
hsa-miR-497-3p
hsa-miR-6715a-3p
Human
Human
hsa-miR-6776-3p
hsa-miR-6715b-3p
hsa-miR-1255b-2-3p
hsa-miR-548
hsa-miR-548
hsa-miR-6793-5p
hsa-miR-421
hsa-miR-371b-5p
hsa-miR-1323
hsa-miR-323a-3p
hsa-miR-4633
hsa-miR-1305
hsa-miR-586
hsa-miR-129-5p
hsa-miR-6504-3p
hsa-miR-491
hsa-miR-4719
hsa-miR-33a
hsa-miR-675-3p
hsa-miR-373-5p
hsa-miR-153-5p
hsa-miR-3671
hsa-miR-3202
hsa-miR-33b
hsa-miR-5680
hsa-miR-4742-3p
hsa-miR-513b-5p
hsa-miR-6715a-3p
hsa-miR-6759-5p
hsa-miR-543
hsa-miR-6748-5p
hsa-miR-23c
hsa-miR-7159-5p
hsa-miR-23a-3p
hsa-miR-23b-3p
hsa-miR-513b
hsa-miR-125b-2-3p
hsa-miR-4457
hsa-miR-499b-3p
hsa-miR-302a-5p

The following optimisation (A4.1-4.3) was performed by Kathryn Dickson (Arkell laboratory member) and has been replicated from her Honours Thesis with permission.

A4.1 HEK293T Reference Gene Selection

To accurately measure the mRNA decay levels of *ZIC2* transcripts *in vitro*, it is important to optimise the RT-qPCR assay before quantification can begin. This requires the selection of reference genes that are expressed in a similar manner to *ZIC2*. Reference genes should have minimal variation in expression in the cell or tissue type examined and be expressed at a similar level to *ZIC2* to ensure there is adequate overlap between their amplification levels. Similar dynamic ranges between *ZIC2* and reference genes will result in accurate quantification of starting transcript copy numbers from the same amount of RNA.

To select appropriate reference genes, the RefGenes tool on the Genevestigator database was utilised (Hruz *et al.*, 2011). Based on the pool of public microarray data available in RefGenes, four reference genes were selected as candidates to be used in the *ZIC2* mRNA assay. These genes were all stably expressed in HEK cells within a similar range to *ZIC2*, and all genes had previously been used in other RT-qPCR assays: *PPP1R8, ABCE1, MAD2L2* and *USP39*. The PrimerBank database (Spandidos *et al.*, 2009) was used to select oligonucleotides for each reference gene that have been successfully used by other researchers in RT-qPCR assays.

A4.2 Reference Gene Efficiency Test

The performance of RT-qPCR assays can be assessed via amplification efficiency and the limit of the dynamic range. The amplification efficiency of a qPCR assay can be affected by many variables including primer design, amplicon size and the presence of PCR inhibitors within the reaction mix.

Before mRNA decay rates of *ZIC2* can be analysed, the RT-qPCR protocol and reference genes must be optimised. The efficiency of the primers chosen to detect mRNA levels can be analysed via the use of a standard curve generated by serial dilutions of the template. A well-optimised assay will produce an efficiency of 90-105%, where 100% represents a doubling of the PCR product every cycle (Bustin *et al.*, 2009). A linear standard curve with a regression line (R^2) of 0.98 or higher also indicates an optimised reaction. To create standard curves for the *ZIC2* and reference gene assays, a $\log_{10}$ dilution series of total RNA template (extracted from untransfected HEK 293T cells) were used, ranging from 100 ng to 0.01 ng per reaction. For each concentration, the mean CT value was plotted against the log of the dilution factor. The efficiencies of the assays were then calculated based on the slope of the linear least square

regression line. The standard curve was plotted using Microsoft Excel and the linear regression and coefficient of determination (R^2) of the standard curve were calculated using the Microsoft Excel LINEST function. The efficiency of the assay was calculated from the slope of the regression line of the standard curve using the formula $E = 10^{-1/slope}$.

The efficiencies for *ZIC2, ABCE1, PPP1R8* and *USP39* were all found to be within the acceptable range (Table 2.4). Additionally, the high regressions (R^2) obtained from the standard curves suggest that each primer set has a dynamic range over at least five orders of magnitude. Thus, these sets of primers are optimised to measure mRNA levels and were therefore employed in the relative quantification of *ZIC2* expression.

A4.3 DRB cell viability assay

Trypan Blue staining was used to assess the viability of HEK 293T cells treated with 60 µM to 300 µM of the transcriptional inhibitor DRB suspended in DMSO. DRB inhibits the kinases responsible for phosphorylating the carboxy-terminal domain of the largest RNA Pol II subunit preventing elongation of transcripts (Bensaude, 2011). Nonviable cells have disrupted membranes allowing the uptake of Trypan Blue, turning the cytoplasm blue, while the intact membranes of viable cells prevents this, retaining a clear cytoplasm (Strober, 2001). The proportion of viable to nonviable cells can be determined visually based on the colour of the cells. Cell suspensions were obtained by detaching adherent cells with Trypsin-EDTA (0.5 g/L; Life Technologies), neutralized in an equal volume of supplemented DMEM as described in section 2.4.1. An equal volume of Trypan Blue stain (0.4% Trypan Blue [Matheson, Coleman and Bell] in 1 X PBS [Amresco]) was added to the cell suspension and gently mixed. A 10 µL aliquot of the cell/Trypan Blue mix was loaded onto a clean haemocytometer and visualised under a light microscope at 100 X magnification. The number of viable and nonviable cells in 4 x 1 mm^2 squares of the haemocytometer grid were counted and used to calculate the proportion of nonviable cells in the suspension.

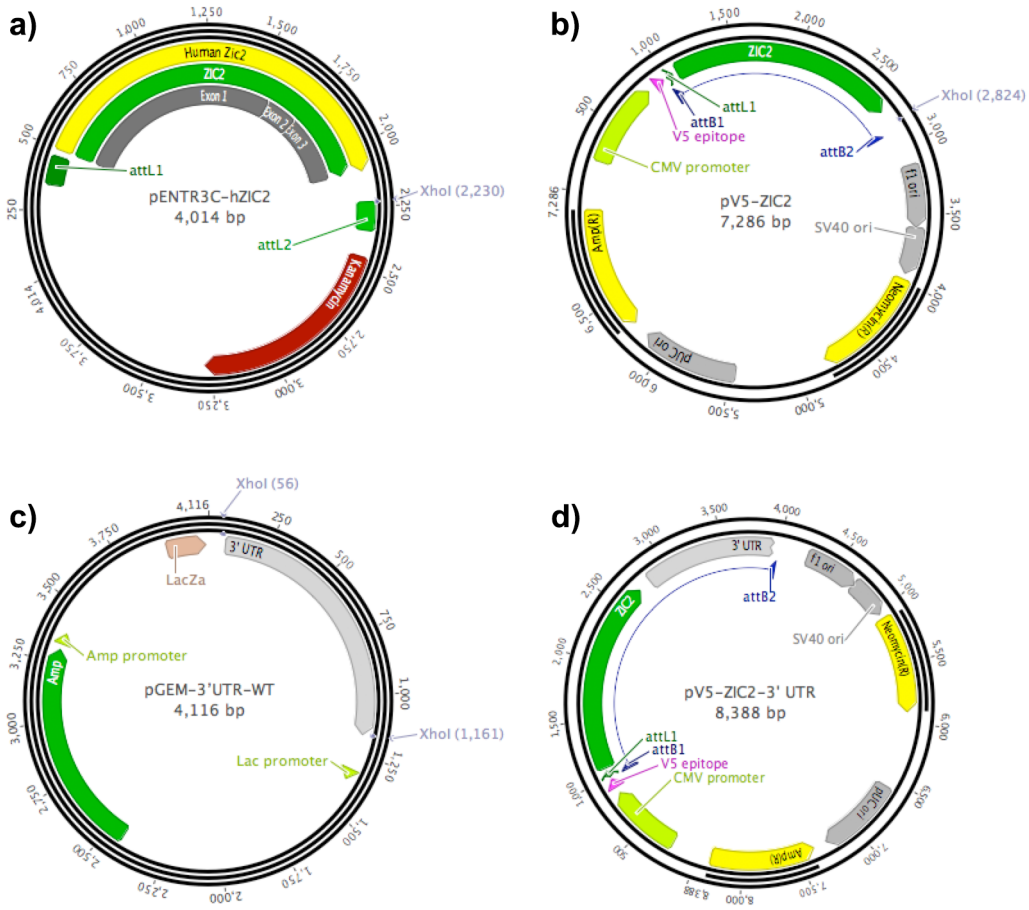


Figure A4.3: Plasmid maps. (a) pENTR3C-h*ZIC2*, **(b)** pV5-*ZIC2,* **(c)** pGEM-3'UTR-WT, **(d)** pV5-ZIC2-3'UTR.

A4.4 Validation of Reference Genes for RT-qPCR Studies of Murine Gastrulation

The following section is taken from a manuscript being prepared for submission. Figures and tables are located at the end of the section.

Kristen S. Barratt, Koula Diamand, Ruth M. Arkell*

Introduction

Early-mid embryogenesis is a period of dynamic expression, characterized by rapid changes in cell division, morphology and number. Pre-fertilisation (3.5 dpc; days post coitum), the blastocyst is made of 32 cells dividing roughly every 12 hours. By the onset of gastrulation (6.5 dpc), the embryo consists of ~900 cells. Those within the proliferative zone of the embryo proper divide approximately every 2-3 hours whilst those outside of this zone divide every ~6.5 hours (Snow 1977). Transcriptome mining of murine embryos revealed genes involved in cellular metabolism and RNA transcription are upregulated at this stage, reflecting a requirement for the acceleration of cell proliferation (Kojima et al. 2014). This rapid turnover of cells can be attributed to the need for a constant source of progenitor cells for germ layer formation (Power and Tam 1993), the specialisation of tissues within the embryo, and multiple signalling pathways, such as TGF superfamily, WNT and FGF, working to establish the major embryonic axes and direct cell movements in a spatially and temporally restricted manner (Pfister et al. 2007). By 7.5 dpc the embryo has expanded to ~15000 cells, with the proliferative zone estimated to produce half of the existing cells at this stage (Snow 1977). Pluripotency genes are downregulated whilst WNT pathway and neurulation genes are upregulated (Kojima et al. 2014). As the specialisation of organs and tissues progresses, organisms become vastly heterogeneous and gene expression becomes compartmentalised and specific. As such, changes in gene expression often occur rapidly and transiently during this time. For example, whilst microarray analysis of murine embryos detected 2-4 fold fluctuations in gene expression between 6.25 and 9.0 dpc, the largest fluctuations were found to occur at the onset of patterning (8.0-8.5 dpc), and were not replicated in the times preceding or following this stage (Mitiku and Baker 2007). By 8.5 dpc, the embryo has doubled in size compared to 7.5 dpc. Here, neurulation genes maintain their high level of expression, accompanied by enrichment of vasculogenesis and organogenesis genes, an increase in transcript diversity, and the downregulation of cell cycling and division genes. This stall in cell proliferation is attributed to cells committing to lineage specification (Mitiku and Baker 2007; Kojima et al. 2014), allowing organogenesis to proceed at 9.5 dpc.

It is these vast fluctuations in cell number and gene expression that create difficulties when attempting to quantify gene expression at gastrulation. While whole-mount in situ hybridization provides spatial gene expression data, it does not possess the ability to quantify this expression. Comparatively, transcriptome analysis provides temporal gene expression analysis, but provides large amounts of data for analysis and cannot be performed without high cost and effort. RT-qPCR, however, affords the user the ability to quantify the small and rapid transcriptional changes that occur during gastrulation and between samples, and is frequently used as a phenotyping tool at these stages to detect gene expression changes in mutant embryos.

In RT-qPCR, the expression level of a target gene is measured against the expression of a stably expressed 'reference gene' to minimise error. Stability in this context refers to the balance between the decay rate of already transcribed mRNA and the rate of gene transcription, resulting in a steady state level of available mRNA. The assumptions made when choosing stable reference genes are that target and reference genes are expressed at a similar level, that expression of the reference gene will be consistent between samples tested, and that it will undergo the same errors as the target during cDNA preparation and subsequent analysis (Vandesompele et al. 2002; Bustin et al. 2009; Guenin et al. 2009; Chapman and Waldenström 2015). In reality, the expression of any reference gene can vary considerably between experimental conditions (Guenin et al. 2009). Previous publications have found that the best candidate reference gene was context dependent and changed between the different tissues being examined. Additionally, as reference genes are considered 'housekeeping genes', their expression is generally strong, which can result in large discrepancies in transcript abundance relative to the target being analysed (Hruz et al. 2011). These issues are compounded when you consider that it is total mRNA being evaluated, and enzymatic reactions (such as the rate of reverse transcription or PCR) can introduce variations that effect the strength of the detected signals. By not addressing these factors, significant biases can be introduced into an experiment, resulting in incorrect data interpretation. As there is no 'one size fits all' reference gene in experimental data, multiple studies, including the MIQE guidelines for the minimum information for publication of RT-qPCR data, recommend using two-to-three concurrent reference genes to validate each data set (Vandesompele et al. 2002; Bustin et al. 2009; Kozera and Rapacz 2013). A recent study, however, identified that these guidelines are not routinely followed, with an average of 1.23 reference genes used in normalisation in published papers (Chapman and Waldenström 2015). To enable the informed selection of reference genes, algorithms such as geNorm (Vandesompele et al. 2002; Mestdagh et al. 2009), Bestkeeper (Pfaffl et al. 2004) and Normfinder (Andersen et al. 2004) were created to determine the most stably expressed referenced genes that can then be used in quantification of targets.

Previous studies by Willems et al (2006) and Veazey et al (2011) have attempted to address these issues by recommending stable reference genes suitable for comparisons of embryonic and extra-embryonic stem cells (Veazey and Golding 2011), or reference genes that are stable from pre-implantation (3.5 dpc), through gastrulation (7.5 dpc) and patterning (9.5 dpc) to organogenesis (11.5 dpc) (Willems et al. 2006). As the recommendations by Willems et al (2006) are stable over such a wide range of embryonic stages, they may not be suitable for specific, restricted periods of time such as gastrulation and patterning (6.5-9.5 dpc). We therefore set out to provide such recommendations.

In this study we compare the expression profiles of five common reference genes (*Actb, Gapdh, H2afz, Tbp* and *Ubc*) to determine those that are most stably expressed during murine gastrulation and thus are most suited for use as normalisation controls in RT-qPCR. To evaluate the stability of candidate genes, total RNA was isolated from embryos at 6.5, 7.5, 8.5 and 9.5 dpc. Upon quantification, RT-qPCR was used to measure the transcript levels of all five candidate reference genes. Via the use of NormFinder, geNorm and BestKeeper algorithms, H2afz and Ubc were deemed the most stably expressed of the five candidates during murine gastrulation. Contingent on this, the expression of *Zic1* and *Zic2* at gastrulation were normalised against H2afz and Ubc, with the resulting expression consistent with previously published expression patterns. Our results suggest that the careful selection of stable reference genes for data normalisation will produce a more accurate quantification of target transcript levels during gastrulation.

Methods

Embryo collection and RNA extraction

Wildtype C3H/HeH mice were maintained in a light cycle of 12 h light: 12 h dark, the midpoint of the dark cycle being 12 A.M. For the production of staged embryos, 12 P.M. on the day of the appearance of the vaginal plug is designated 0.5 dpc. Embryos were collected at 6.5, 7.5, 8.5 and 9.5 dpc, staged according to Downs and Davies (Downs and Davies 1993), and pooled based on their age. The total RNA was extracted immediately with the NucleoSpin® RNA kit (Macherey-Nagel) according to the manufacturer's instructions and quantified via nanodrop. Additional DNaseI treatment (#04716728001, 10U, Roche) was performed to ensure no genomic DNA contamination was present. This was confirmed via standard PCR with oligonucleotides that spanned intron 1 of *Shh* (RA#748: CAC ACA CAC ATT TCT CTG TCC and RA#749: TAG CTC AGT GCT TGC AAG GTT A), whereby no amplification indicated the samples were free from genomic DNA. RNA integrity was confirmed via agarose gel electrophoresis. RNA samples were stored for up to two weeks maximum at -80°C before RT-qPCR analysis.

Primer design and RT-qPCR analysis

Primer sequences for *Actb* and *Ubc* were obtained from Mamo et al (2007), whilst sequences for *Gapdh, H2afz* and *Tbp* were obtained from Primer Bank (Wang and Seed 2003). Primers were selected over intron-exon boundaries whenever possible to limit genomic DNA amplification.

One-step RT-qPCR was performed on 50 ng of RNA per sample, using the SensiFAST SYBR Hi-ROX One-Step Kit (Bioline) at the recommended conditions. cDNA conversion and melt curve analysis were performed using a StepOnePlus™ Real-Time PCR System (Applied Biosystems®) and StepOne software (version 2.2.2; Applied Biosystems®). PCR efficiencies were calculated using the slope of a calibration curve as described in Bustin et al (2009). For quantification, the assay for each gene consisted of three internal replicates per gene per embryo age. At least three independent experiments were performed for each RT-qPCR. Mean values were calculated and the C_T value used for analysis with Bestkeeper, converted into a logarithmic scale via the calculation 2^{Cq} for analysis with Normfinder or a linear scale for analysis with geNorm. The relative fold expression of Zic1 and Zic2 were calculated as per the Livak method (Livak and Schmittgen 2001).

Results

A survey of the literature identified five commonly used candidate reference genes in mouse embryonic studies - *Actb, Gapdh, H2afz, Tbp* and *Ubc* (Table A4.1). To determine which of these candidates was most stable during mouse gastrulation, total RNA was extracted and purified from pooled wildtype embryos at 6.5, 7.5, 8.5 and 9.5 dpc (n=3 independent pools per age, with n=4+ embryos per pool). The RNA was quantified and used to determine both the efficiency of the primers in the RT-qPCR reaction, as well as the stability of the candidate reference genes. The final C_T values were analysed for stability using Normfinder, geNorm and Bestkeeper.

Primer efficiency

All of the reference gene primer pairs were tested for their amplification efficiency. A primer pair was considered efficient if they produced an R2 value ≥0.98 and, ideally, a slope value of -3.32 resulting in an efficiency of 100% (where 100% represents doubling of the PCR product every cycle). Due to normal experimental variability, efficiency values between 90-110% were also acceptable for the conditions of this RT-qPCR assay. The efficiencies of the RT-qPCR assays for *Actb, Gapdh, Hprt, H2afz, Tbp, Ubc, Zic1* and *Zic2* were all found to be within the accepted range (n=3, Table A4.2) and thus could be used to analyse the stability of the candidate reference genes.

Analysis of Reference Gene Stability

Normfinder

Mean CT values for each reference gene at each embryonic stage were used to calculate an RQ value, which could be analysed with Normfinder (Andersen et al. 2004). By taking into account inter- and intra- group variation, a stability value is calculated. The most stably expressed candidate gene during gastrulation was calculated to be *Gapdh* (stability value 0.212) (Table A4.3). *Ubc* (0.247) and *H2afz* (0.251) were comparatively unstable on their own and therefore would not be good choices for normalisation of a target individually. Normfinder does, however, predict that a combination of both *Ubc* and *H2afz* (with a joint stability value of 0.151) will provide the most consistent results when normalising a target, rather than using *Gapdh* in conjunction with *Ubc*.

geNorm

Mean CT values for each reference gene at each embryonic stage were used to calculate an M value and CV value via geNorm. The M value represents the average pairwise variation of a specific gene in comparison to all other genes, with a low M value (<1) signifying the most stable gene in a heterogeneous sample (Hellemans et al. 2007; Guenin et al. 2009). Similarly, the CV value indicates the intra-assay coefficient of variation, with a low number (<0.5) representing stability (Hellemans et al. 2007). Comparable to the results produced by Normfinder, geNorm predicted *H2afz* to be the most stably expressed reference gene tested, followed by *Ubc* (Table A4.3).

Bestkeeper

Mean CT values for each reference gene at each embryonic stage were analysed with Bestkeeper (Pfaffl et al. 2004) and an R value generated. The R value represents the Pearson coefficient of correlation, with a value of or close to 1 representing a stable gene. Similar to the results obtained by geNorm, Bestkeeper determined *H2afz* to be the most stable reference gene (Table 3). Gapdh was also determined to be relatively stable, followed by *Ubc* and *Tbp*. Comparatively, *Actb* was considered the most unstable reference gene for all three algorithms tested.

As *H2afz* and *Ubc* were predicted to be the most stable combination of genes by Normfinder with a joint stability value of 0.151, and both genes were ranked 1 and 2 respectively in GeNorm, and 1 and 3 respectively in Bestkeeper, they were chosen as the most stable of the tested reference genes to normalise further gene expression studies to. It should be noted that Gapdh performed comparatively well to *Ubc*, and thus could act as a sufficient substitute in future

analysis. *Actb* and *Tbp*, however, were consistently ranked the least stable of the genes tested and are thus not recommended for use in further gene expression studies which span gastrulation. Taken together, the data from all three algorithms predict *Ubc* and *H2afz* as stable reference genes and are therefore recommended for RT-qPCR normalisation during mouse gastrulation.

Zic expression during gastrulation

The expression of the transcription factors *Zic1* and *Zic2* were analysed using the reference genes deemed to be most stable during this period of time – *Ubc* and *H2afz*. *Zic2* is known to be highly active during axis formation and node development at 6-7.5 dpc (Elms et al. 2004; Warr et al. 2008), as well as neural crest induction at 8-8.5 dpc (Elms et al. 2003), with *Zic2* loss-of-function resulting in congenital malformations such as Holoprosencephaly (Brown et al. 2005) and cardiac defects reminiscent of Heterotaxy (Barratt et al. 2014). In comparison, the *Zic* family member *Zic1* exhibits no detectable expression prior to 8.5 dpc (Elms et al. 2004). As neurulation progresses, *Zic1* is first reliably detected in the trunk neural tube in a similar pattern to *Zic2*, with both genes later expressed in the dorsal spinal cord, dorsal cranial neural tube and the somites (Nagai et al. 1997; Elms et al. 2004). As such, the comparative expression of *Zic1* and *Zic2* throughout gastrulation and organogenesis provide an opportunity to test the effectiveness of *H2afz* and *Ubc* as stable reference genes.

When normalised to either *Ubc* or *H2afz* using the $2^{-\Delta\Delta CT}$ method (Figure A4.4), the relative fold expression of *Zic1* was relatively undetectable from 6.5-8.5 dpc and peaked at 9.5 dpc, consistent with the known expression pattern of *Zic1* mRNA during embryogenesis. Comparatively, the relative fold expression of *Zic2* was found to peak at 6.5 dpc when normalised to either reference gene, demonstrating a gradual decrease in expression by 9.5 dpc. This reduction in expression levels is presumably due to changes in the ratio of *Zic2* expressing tissue compared non-*Zic2* expressing tissues at these later stages. Whilst *Zic2* expression at 6.5 dpc encompasses the majority of the embryo (being expressed in the extra embryonic and embryonic ectoderm of the egg cylinder [Elms et al. 2004]), by 9.5 dpc expression is tissue specific and restricted to only the dorsal portion of the embryo.

Discussion

The selection of a reference gene that is stable throughout the tissues and time periods being examined is critical for correct RT-qPCR analysis. Additionally, the use of multiple reference genes to analyse the same data is essential to prevent misinterpretation of result. The dynamic changes that occur in embryonic tissues during development pose a particular challenge when

attempting to analyse gene expression at specific stages. In this study, we evaluated and compared the stability of five candidate reference genes during gastrulation (6.5 dpc) to the start of organogenesis (9.5 dpc) via the use of RT-qPCR in combination with the statistical tools geNorm, Normfinder and Bestkeeper. Past studies have found that there is no large difference between the output rankings between each algorithm (Willems et al. 2006; De Spiegelaere et al. 2015). A similar result was found in this study, with *Ubc* and *H2afz* consistently performing better when analysed for stability with each tool. Gapdh was also considered stable by two of the three analysis tools utilised, and therefore remains a viable candidate when analysing gene expression at gastrulation.

Previous analysis of stable reference genes spanning 3.5-9.5 dpc by Willems et. al. identified Actb and Tbp as the most stably expressed reference genes (Willems et al. 2006). This result contrasts with our analysis presented here, where Actb and Tbp consistently ranked the least stable in each of the three algorithms used. This difference in recommendation from both studies can presumably be attributed to the wider range of embryonic stages covered by Willems et al, suggesting that whilst *Actb* and *Tbp* are stable when comparing large changes in embryonic development stage, they are not stably expressed during the restricted period of gastrulation. The contrasting results presented here, and those contrasting results achieved by other studies (Meller et al. 2005; Zhang et al. 2005; Hruz et al. 2011), reflect the importance of ensuring the reference genes used in a study are suitable to the tissue stage and type being examined.

Following the selection of *Ubc* and *H2afz* as stable reference genes, their validity was tested via the relative quantification of *Zic1* and *Zic2* expression at gastrulation and early organogenesis. The results achieved in this paper correlate with the known expression profiles of both genes during this time period, and past published RT-PCR results (Elms et al. 2004; Houtmeyers et al. 2013) whereby *Zic2* expression remains high throughout gastrulation whilst Zic1 expression is relatively undetectable until cranial neural crest development at 8.5 dpc.

As technologies progress and large-scale data analysis becomes the norm, stable reference genes are routinely being selected from within microarray experiments. These reference genes routinely perform better than those selected from external RT-qPCR studies, owing to the fact that the microarray reference genes are already exposed to the same conditions as the genes being targeted for quantification. Errors still exist in these methods, however, with probe choice remaining an issue due to incomplete hybridization and non-specific binding (Hruz et al. 2011). Though some databases have complied this microarray data for aid in reference gene selection,

they still do not currently have the specificity to identify stable genes spanning murine gastrulation time points, despite having over 50,000 experimentally validated microarray entries (Hruz et al. 2011). Similarly, whilst new tools such as iTranscriptome can be used to establish whether a candidate reference gene is expressed in high levels in a 7.0 dpc mouse embryo (for example, *H2afz* is expressed ten-fold higher than *Ubc* at this stage) via a digital WMISH (Peng et al. 2016), it cannot yet provide information about the expression of candidates across a range of time points. Until such time as these technologies progress, the evaluation of candidate reference genes via stability assays are essential before target RT-qPCR quantification can occur.

Table A4.1: Description of the six candidate reference genes analysed.

Symbol	Name	Reference Sequence
Actb	Beta-actin	NM_007393
Gapdh	Glyceraldehyde-3-phosphate dehydrogenase	GU214026
H2afz	H2A histone family, member Z	NM_016750
Tbp	TATA box binding protein	U63933
Ubc	Ubiquitin	NM_019639

Table A4.2: Primer sequences and their mean RT-qPCR efficiencies (n=3). An efficiency of 100% represents doubling of the PCR product every cycle. A primer pair was considered efficient if R^2 value ≥0.98 and an efficiency value between 90-110%.

Gene		Primer sequence (5' – 3')	Primer location	Product size (bp)	Slope	R^2	Efficiency
Actb	F	ATG AGC TGC CTG ACG GCC AGG TCA TC	Exon 3	192	-3.34	0.99	99.17%
	R	TGG TAC CAC CAG ACA GCA CTG TGT TG	Exon 4				
Gapdh	F	TGA CGT GCC GCC TGG AGA AA	Exon 4	98	-3.28	0.99	101.81%
	R	AGT GTA GCC CAA GAT GCC CTT CAG	Exon 5				
H2afz	F	GCG CAG CCA TCC TGG AGT A	Exon 3	202	-3.30	0.99	101.00%
	R	CCG ATC AGC GAT TTG TGG A	Exon 5				
Tbp	F	GAA GAA CAA TCC AGA CTA GCA GCA	Exon 5	128	-3.36	0.99	98.29%
	R	CTT ATG GGG AAC TTC ACA TCA CAG	Exon 5				
Ubc	F	CGT CGA GCC CAG TGT TAC CAC CAA GAA GG	Exon 1	112	-3.33	0.99	99.47%
	R	CCC CCA TCA CAC CCA AGA ACA AGC ACA AG	Exon 1				
Zic2	F	TCG TTG CGG AAG CAC ATG AA	Exon 2	178	-3.27	0.98	102.34%
	R	ACA GGT TGG AGC TGC TTT GT	Exon 3				
Zic1	F	GCA AGA TGT GCG ATA AGT CC	Exon 2	159	-3.28	0.99	101.84%
	R	GGT TGT CTG TTG TGG GAG AC	Exon 3				

Table A4.3: Stability ranking of the five candidate reference genes using Normfinder, geNorm and Bestkeeper. For Normfinder, a low stability value indicates a stably expressed gene (Andersen *et al.*, 2004), whilst for Bestkeeper, a ranking close or equal to 1 indicates the most stable gene (Pfaffl *et al.*, 2004). For geNorm, M value of <1 and a CV value of <0.5 is recommended for heterogeneous samples (Hellemans *et al.*, 2007).

Stability Ranking	Normfinder		geNorm			Bestkeeper	
	Gene	Stability Value	Gene	M value	CV value	Gene	R Value
1	*Gapdh*	0.212	*H2afz*	0.540	0.095	*H2afz*	0.999
2	*Ubc**	0.247	*Ubc*	0.674	0.247	*Gapdh*	0.986
3	*H2afz**	0.251	*Gapdh*	0.770	0.311	*Ubc*	0.775
4	*Tbp*	0.254	*Tbp*	0.812	0.351	*Tbp*	0.773
5	*Actb*	0.403	*Actb*	0.832	0.374	*Actb*	0.520

* Whilst *Gapdh* is calculated to be the most individual stable gene by Normfinder, *Ubc* and *H2afz* are recommended as the best combination of two genes with a joint stability value of 0.151.

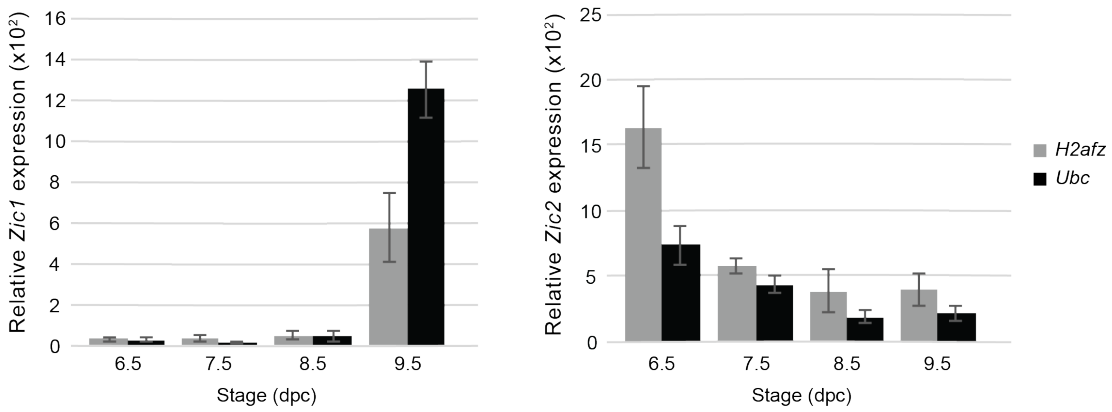


Figure A4.4: Relative *Zic1* and *Zic2* expression during murine gastrulation, normalised to *Ubc* and *H2afz*. Expression was calculated using the $2^{-\Delta\Delta CT}$ method.

Paper references

Andersen, C.L., Jensen, J.L., and Ørntoft, T.F., 2004. Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and Colon Cancer Data Sets. Cancer Research 64(15), 5245–5250.

Barratt, K.S., Glanville-Jones, H.C., and Arkell, R.M., 2014. The Zic2 gene directs the formation and function of node cilia to control cardiac situs. Genesis 52(6), 626–635.

Brown, L.Y., Paraso, M., Arkell, R.M., and Brown, S., 2005. In vitro analysis of partial loss-of-function ZIC2 mutations in holoprosencephaly: Alanine tract expansion modulates DNA binding and transactivation. Human Molecular Genetics 14(3), 411–420.

Bustin, S., Benes, V., and Garson, J., 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. Clinical Chemistry 55(4), 611–22.

Chapman, J.R., and Waldenström, J., 2015. With Reference to Reference Genes: A Systematic Review of Endogenous Controls in Gene Expression Studies. PLoS ONE 10(11), e0141853.

De Spiegelaere, W., Dern-Wieloch, J., Weigel, R., Schumacher, V., Schorle, H., Nettersheim, D., Bergmann, M., Brehm, R., Kliesch, S., Vandekerckhove, L., and Fink, C., 2015. Reference gene validation for RT-qPCR, a note on different available software packages. PLoS ONE 10(3), e0122515.

Downs, K.M., and Davies, T., 1993. Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. Development 118(4), 1255–1266.

Elms, P., Scurry, A., Davies, J., Willoughby, C., Hacker, T., Bogani, D., and Arkell, R.M., 2004. Overlapping and distinct expression domains of Zic2 and Zic3 during mouse gastrulation. Gene Expression Patterns 4(5), 505–511.

Elms, P., Siggers, P., Napper, D., Greenfield, A., and Arkell, R.M., 2003. Zic2 is required for neural crest formation and hindbrain patterning during mouse development. Developmental Biology 264(2), 391–406.

Guenin, S., Mauriat, M., Pelloux, J., Van Wuytswinkel, O., Bellini, C., and Gutierrez, L., 2009. Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. Journal of Experimental Botany 60(2), 487–493.

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., and Vandesompele, J., 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome biology 8(2), R19.

Houtmeyers, R., Souopgui, J., Tejpar, S., and Arkell, R.M., 2013. The ZIC gene family encodes multi-functional proteins essential for patterning and morphogenesis. Cellular and Molecular Life Sciences 70(20), 3791–3811.

Hruz, T., Wyss, M., Docquier, M., Pfaffl, M.W., Masanetz, S., Borghi, L., Verbrugghe, P., Kalaydjieva, L., Bleuler, S., Laule, O., Descombes, P., Gruissem, W., and Zimmermann, P., 2011. RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization. BMC Genomics 12, 156.

Kojima, Y., Tam, O.H., and Tam, P.P.L., 2014. Timing of developmental events in the early mouse embryo. Seminars in Cell & Developmental Biology 34, 65–75.

Kozera, B., and Rapacz, M., 2013. Reference genes in real-time PCR. Journal of applied genetics 54(4), 391–406.

Livak, K.J., and Schmittgen, T.D., 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. Methods 25(4), 402–408.

Meller, M., Vadachkoira, S., Luthy, D.A., and Williams, M.A., 2005. Evaluation of housekeeping genes in placental comparative expression studies. Placenta 26(8–9), 601–607.

Mestdagh, P., Van Vlierberghe, P., De Weer, A., Muth, D., Westermann, F., Speleman, F., and Vandesompele, J., 2009. A novel and universal method for microRNA RT-qPCR data normalization. Genome Biology 10(6), R64.

Mitiku, N., and Baker, J.C., 2007. Genomic Analysis of Gastrulation and Organogenesis in the Mouse. Developmental Cell 13(6), 897–907.

Nagai, T., Aruga, J., Takada, S., Günther, T., Spörle, R., Schughart, K., and Mikoshiba, K., 1997. The expression of the mouse Zic1, Zic2, and Zic3 gene suggests an essential role for Zic genes in body pattern formation. Developmental Biology 182(2), 299–313.

Peng, G., Suo, S., Chen, J., Chen, W., Liu, C., Yu, F., Wang, R., Chen, S., Sun, N., Cui, G., Song, L., Tam, P.P.L., Han, J.-D.J., and Jing, N., 2016. Spatial Transcriptome for the Molecular Annotation of Lineage Fates and Cell Identity in Mid-gastrula Mouse Embryo. Developmental Cell 36(6), 681–697.

Pfaffl, M.W., Tichopad, A., Prgomet, C., and Neuvians, T.P., 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. Biotechnology letters 26(6), 509–15.

Pfister, S., Steiner, K.A., and Tam, P.P.L., 2007. Gene expression pattern and progression of embryogenesis in the immediate post-implantation period of mouse development. Gene expression patterns : GEP 7(5), 558–73.

Power, M.-A., and Tam, P.P.L., 1993. Onset of gastrulation, morphogenesis and somitogenesis in mouse embryos displaying compensatory growth. Anatomy and Embryology 187(5), 493–504.

Snow, M.H.L., 1977. Gastrulation in the mouse: Growth and regionalization of the epiblast. Development 42(1).

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F., 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome biology 3(7), RESEARCH0034.

Veazey, K.J., and Golding, M.C., 2011. Selection of stable reference genes for quantitative rt-PCR comparisons of mouse embryonic and extra-embryonic stem cells. PLoS ONE 6(11), e27592.

Warr, N., Powles-Glover, N., Chappell, A., Robson, J., Norris, D., and Arkell, R.M., 2008. Zic2-associated holoprosencephaly is caused by a transient defect in the organizer region during gastrulation. Human Molecular Genetics 17(19), 2986–96.

Willems, E., Mateizel, I., Kemp, C., Cauffman, G., Sermon, K., and Leyns, L., 2006. Selection of reference genes in mouse embryos and in differentiating human and mouse ES cells. The International journal of developmental biology 50(7), 627–35.

Zhang, X., Ding, L., and Sandford, A.J., 2005. Selection of reference genes for gene expression studies in human neutrophils by real-time PCR. BMC Molecular Biology 6(1), 4.

A4.5 Identification of PAS oligonucleotide non-specific targets

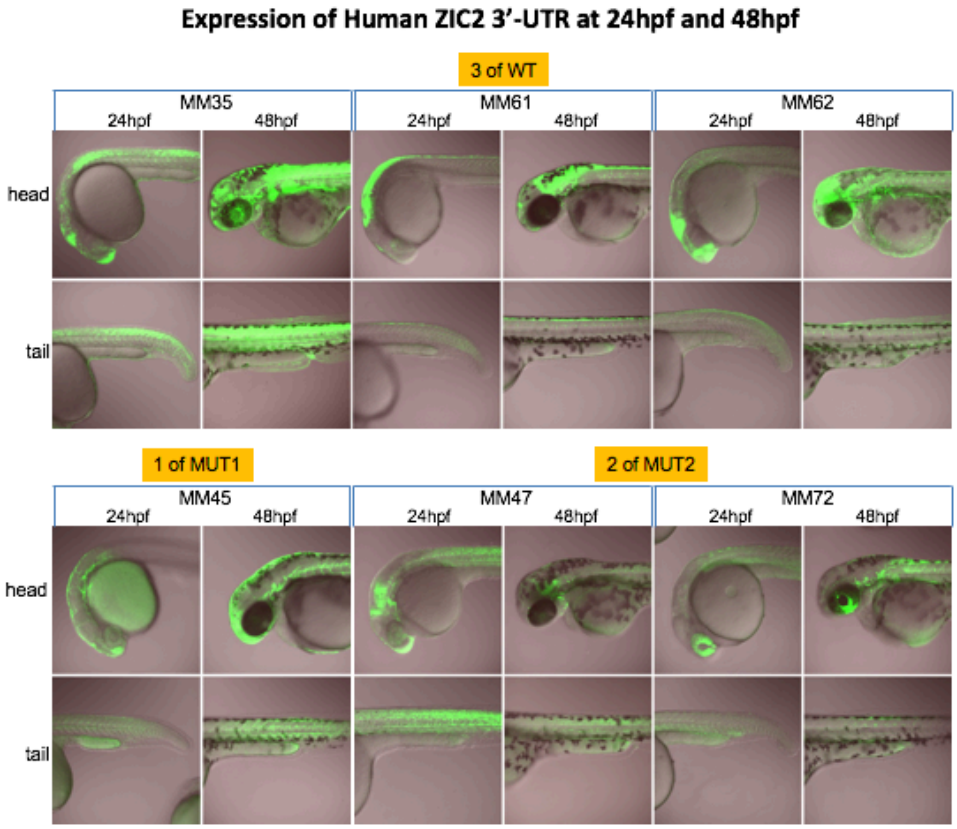
Whilst the addition of betaine into the RT-qPCR amplification reactions reduces the likelihood of primer dimers forming, it does not discount non-specific off-target amplification. Primer-BLAST was utilised to check for off-target oligonucleotide binding for P2 oligonucleotides (Ye *et al.*, 2012). The ten most likely off-target amplicons are listed in Table A4.4. Of the ten possibilities, the first is the *Zic2* 3'UTR, confirming that the oligonucleotide pair is targeting this mRNA. The remaining nine all require >2 nucleotide mismatches between each oligonucleotide and the target DNA for amplification to take place. Four of the non-specific targets are unlikely to be efficiently amplified due to an amplicon length of >500 nt. Additionally, amplification of each non-specific target is unlikely as at least one nucleotide mismatch occurs within the five most 3' nucleotides for each target. It is well documented that mismatches in the 3' end of oligonucleotides affect target amplification significantly more than mismatches in the 5' end. Whilst two nucleotide mismatches have been shown to prevent amplification at the 3' end, a single nucleotide mismatch will result in reduced amplification efficiency (Stadhouders *et al.*, 2010; Ye *et al.*, 2012). As such, it is possible that the increase that is seen at P2 could be due to off-target amplification. It should be noted, however, that each *Zic2* 3'UTR poly(A) oligonucleotide pair is predicted to produce multiple off-target amplicons when run through Primer-BLAST.

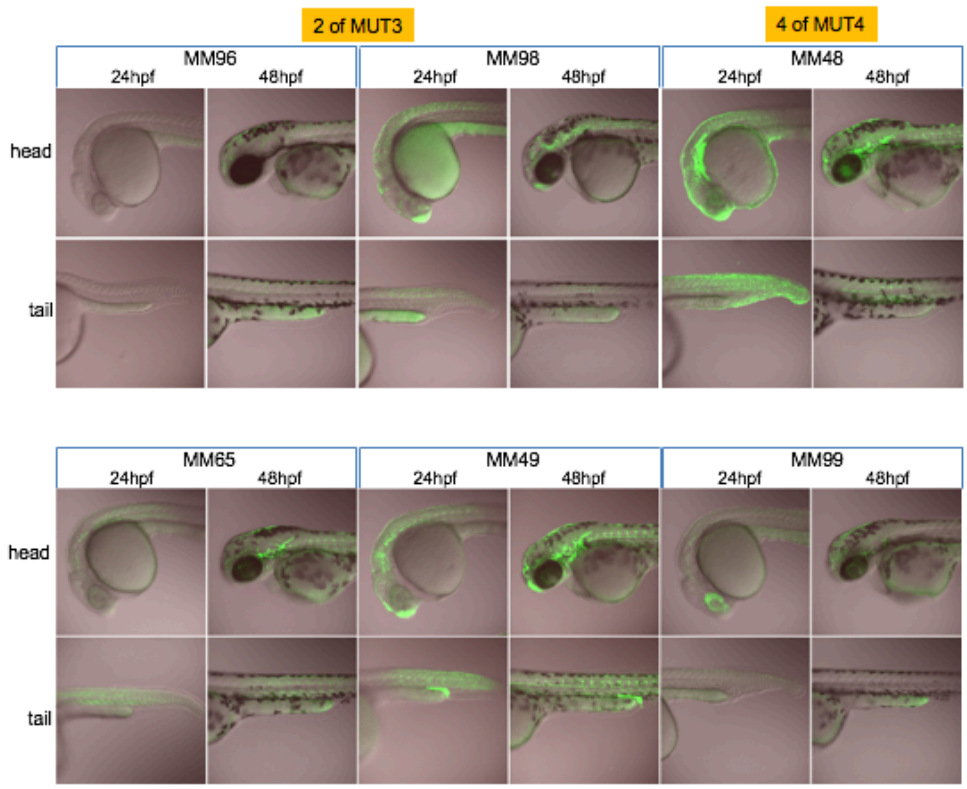
Table A4.4: P2 oligonucleotide off-target amplicons. Off-target amplicons were predicted from Primer-BLAST (Ye et al. 2012). 3' end mismatch: a mismatch occurs between the oligonucleotide and at least one of five most 3'nucleotides in the template. *: amplicon size is too large for efficient amplification under the RT-qPCR conditions used.

Gene	Amplicon length (bp)	Forward primer mismatches	Forward primer mismatches	3' end mismatch
Zic2	101	0	0	0/0
Pcca, transcript variant X1	2438*	3	2	1/1
Clstn3	2222*	4	2	1/0
Tnrc6a, transcript variant X15	104	3	3	1/1
Adgre5, transcript variant X3	169	3	4	1/0
	113	4	5	0/1
Cacna1c	832*	3	4	1/0
Slc35d1	212	4	3	1/0
Htr4	364	3	4	1/1
Slc8a3	3724*	4	3	1/1

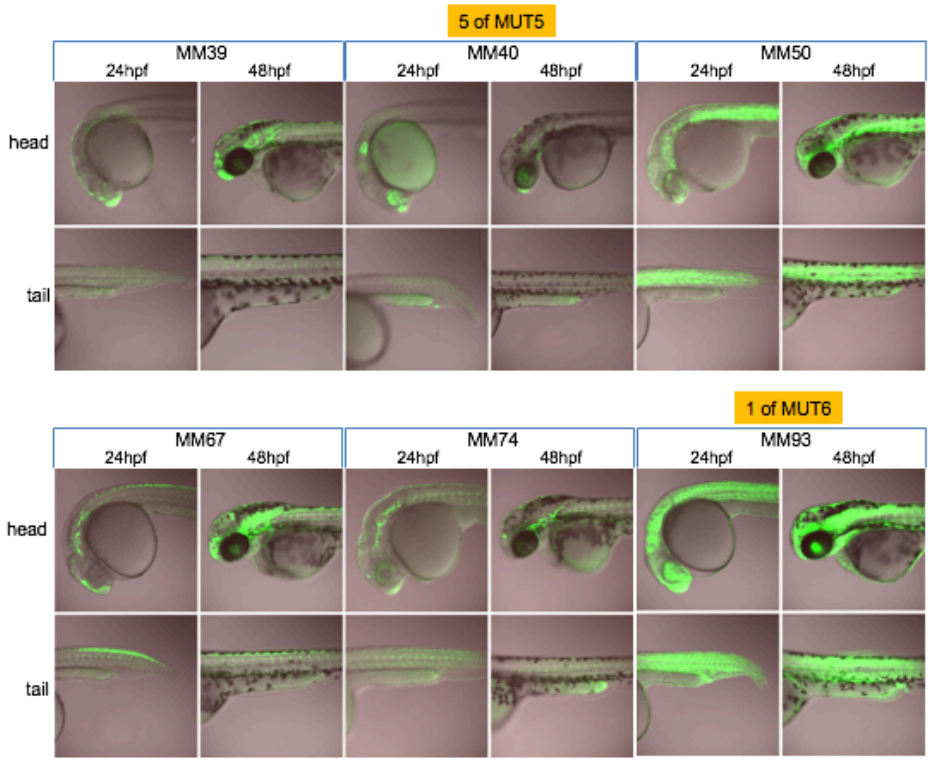
Appendix A5

Figure A5.1: *ZIC2* NCE-GFP transgenic zebrafish exhibited inconsistent reporter expression. Analysis of wildtype (WT) and mutant (MUT1-6) zebrafish expressing GFP under the control of the *ZIC2* NCE. Expression of GFP in founder lines (MM) created with the same 3'UTR constructs was inconsistent; therefore, no definitive expression pattern could be discerned. Photos were provided by S.K. Hong, E. Roessler and M. Muenke at the NHGRI, NIH (Maryland, USA).





Expression of Human ZIC2 3'-UTR at 24hpf and 48hpf



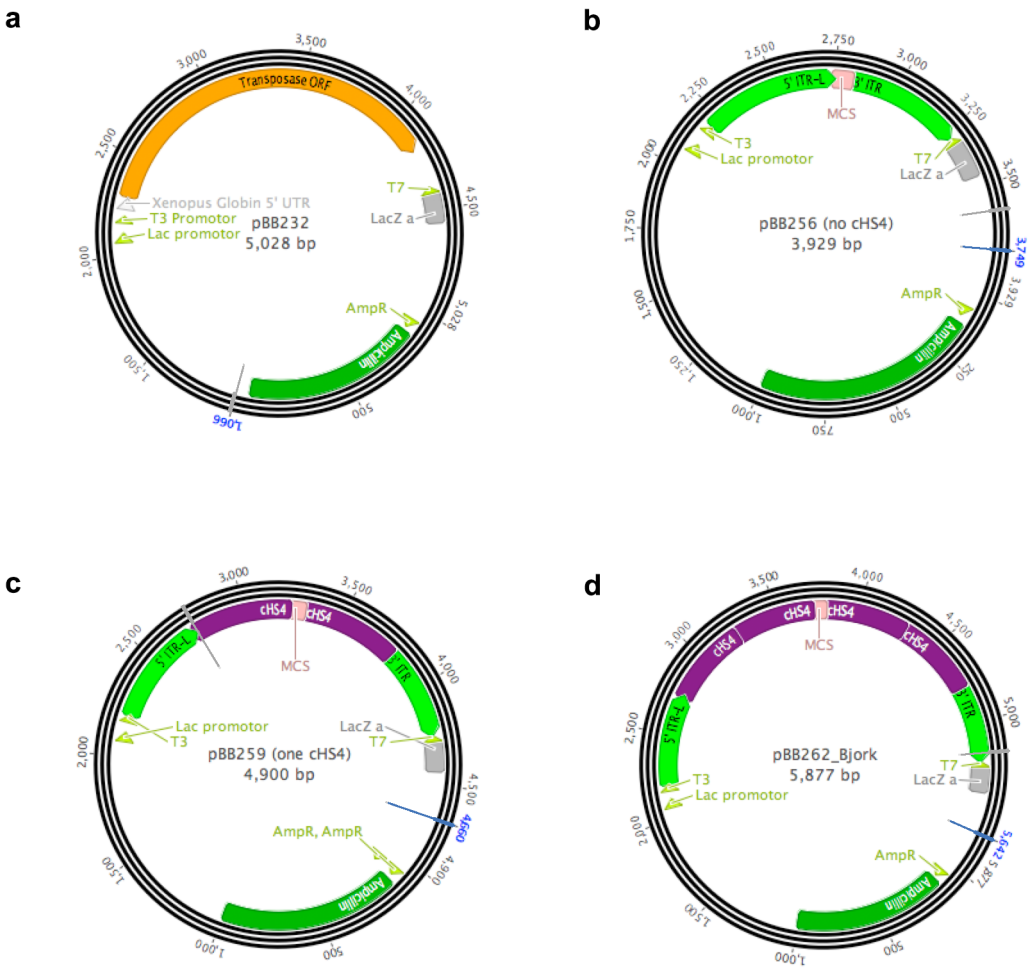
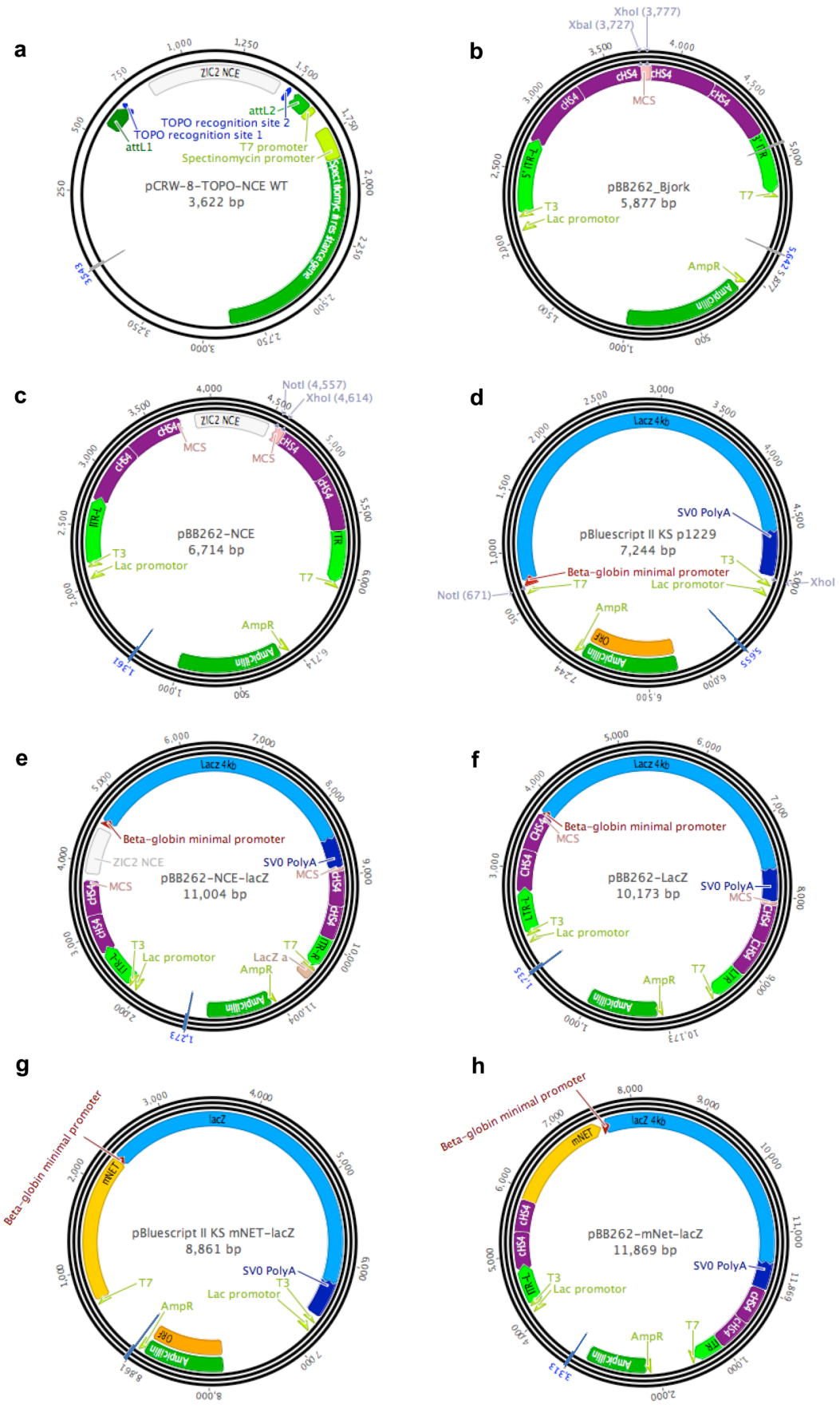


Figure A5.2: Plasmid maps. (a) pBB232, **(b)** pBB256, **(c)** pBB259, **(d)** pBB262. cHS4: Chicken insulator fragments; MCS: multiple coning site; ITR: inverted terminal repeats; ORF: open reading frame.

Figure A5.3: Plasmid maps. (a) pCRW-8-TOPO-NCE, **(b)** pBB262 with *XbaI* and *XhoI* restriction sites annotated, **(c)** pBB262-NCE with *NotI* and *XhoI* restriction sites annotated, **(d)** p1229 with *NotI* and *XhoI* restriction sites annotated, **(e)** pBB262-NCE-*lacZ*, **(f)** pBB262-*lacZ* (negative control). **(g)** pBSK-mNet-*lacZ*, **(h)** pBB262-mNet-*lacZ* (positive control). cHS4: Chicken insulator fragments; MCS: multiple coning site; ITR: inverted terminal repeats; ORF: open reading frame.


a
ZIC2 NCE
attL2
TOPO recognition site 2
TOPO recognition site 1
attL1
T7 promoter
Spectinomycin promoter
pCRW-8-TOPO-NCE WT
3,622 bp
b
XhoI (3,777)
XbaI (3,727)
cHS4
MCS
pBB262_Bjork
5,877 bp
T7
T3
Lac promotor
AmpR
Ampicillin
c
NotI (4,557)
XhoI (4,614)
ZIC2 NCE
MCS
pBB262-NCE
6,714 bp
T3
Lac promotor
T7
AmpR
Ampicillin
d
Lacz 4kb
SV0 PolyA
pBluescript II KS p1229
7,244 bp
Beta-globin minimal promoter
T7
T3
Lac promotor
NotI (671)
XhoI
AmpR
Ampicillin
ORF
e
Lacz 4kb
Beta-globin minimal promoter
ZIC2 NCE
SV0 PolyA
MCS
pBB262-NCE-lacZ
11,004 bp
T3
Lac promotor
T7
LacZ a
AmpR
Ampicillin
f
Lacz 4kb
Beta-globin minimal promoter
MCS
pBB262-LacZ
10,173 bp
SV0 PolyA
T3
Lac promotor
T7
AmpR
Ampicillin
g
Beta-globin minimal promoter
lacZ
mNET
pBluescript II KS mNET-lacZ
8,861 bp
SV0 PolyA
T7
T3
Lac promotor
AmpR
Ampicillin
ORF
h
Beta-globin minimal promoter
lacZ 4kb
mNET
pBB262-mNet-lacZ
11,869 bp
SV0 PolyA
T3
Lac promotor
T7
AmpR
Ampicillin

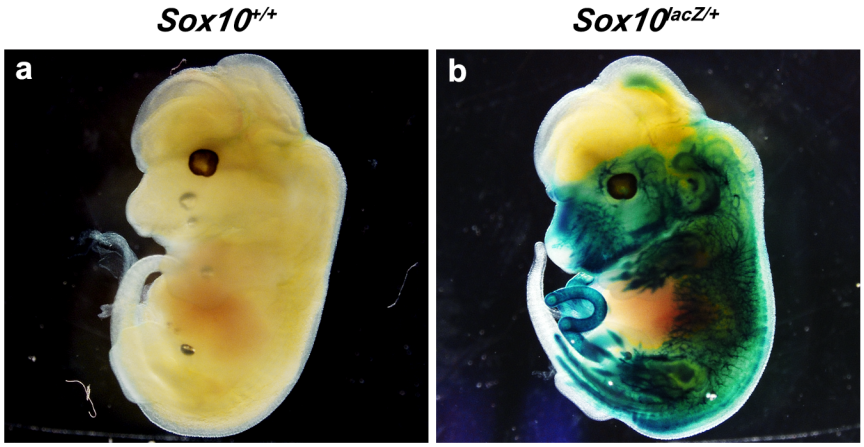


Figure A5.4: *lacZ* expression is detected *Sox10* embryos. (a) *Sox10*$^{+/+}$ and **(b)** *Sox10*$^{(lacZ/+)}$ embryos at 14.5 dpc; anterior is to the left. *Sox10* embryos were used to batch test X-Gal solution.

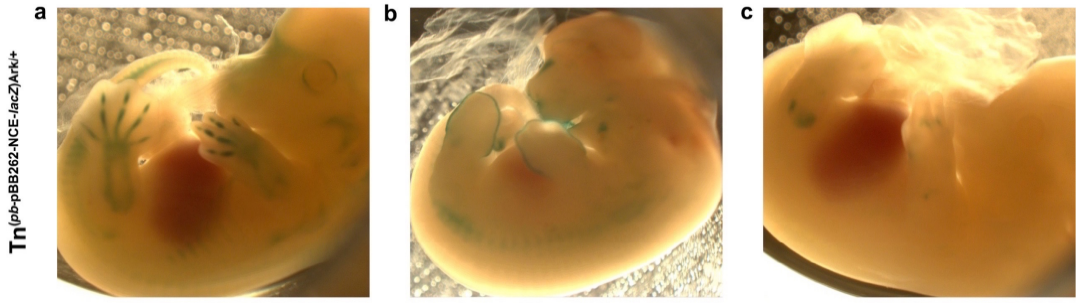


Figure A5.5: *lacZ* expression in Tn$^{(pb\text{-pBB262-NCE-}lacZ)Ark/+}$ pBB262-NCE embryos. (a-c) Tn$^{(pb\text{-pBB262-NCE-}lacZ)Ark/+}$ with varying *lacZ* expression patterns. Expression patterns were inconsistent and unreplicable amongst *lacZ* positive embryos. Photographs taken by Gene Elliot, NIH transgenic Core Facility (Maryland, USA).

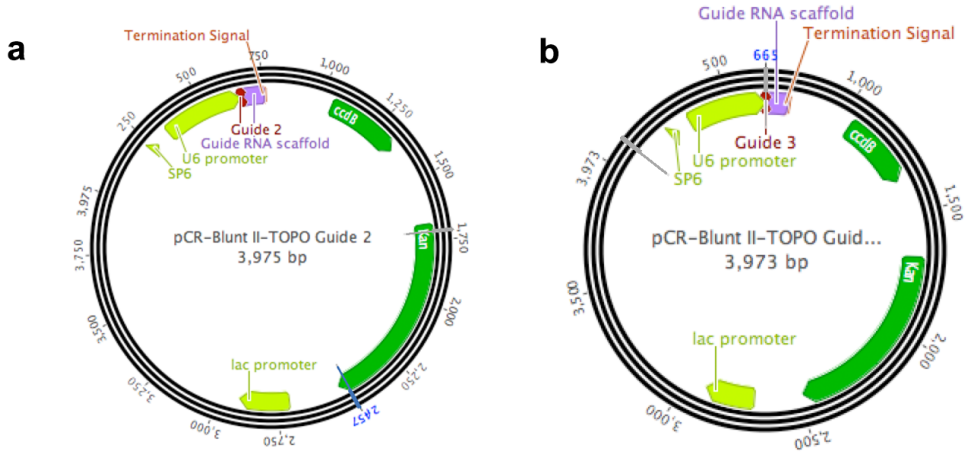


Figure A5.6: CRISPR guide plasmid maps. (a) pCR-Blunt II-TOPO-G2 **(b)** pCR-Blunt II-TOPO-G3.

Abe, Y., Oka, A., Mizuguchi, M., Igarashi, T., Ishikawa, S., Aburatani, H., Yokoyama, S., Asahara, H., Nagao, K., Yamada, M., and Miyashita, T., 2009. EYA4, deleted in a case with middle interhemispheric variant of holoprosencephaly, interacts with SIX3 both physically and functionally. *Human Mutation* **30**(10), 946–955.

Abelson, J.F., Kwan, K.Y., O'Roak, B.J., Baek, D.Y., Stillman, A.A., Morgan, T.M., Mathews, C.A., Pauls, D.L., Rasin, M.-R., Gunel, M., Davis, N.R., Ercan-Sencicek, A.G., Guez, D.H., Spertus, J.A., Leckman, J.F., Dure, L.S., Kurlan, R., Singer, H.S., Gilbert, D.L., Farhi, A., Louvi, A., Lifton, R.P., Sestan, N., and State, M.W., 2005. Sequence variants in SLITRK1 are associated with Tourette's syndrome. *Science* **310**(5746), 317–320.

Adachi, H., Saijoh, Y., Mochida, K., Ohishi, S., Hashiguchi, H., Hirao, A., and Hamada, H., 1999. Determination of left/right asymmetric expression of Nodal by a left side-specific enhancer with sequence similarity to a Lefty-2 enhancer. *Genes & Development* **13**(12), 1589–600.

Affolter, M., Pyrowolakis, G., Weiss, A., and Basler, K., 2008. Signal-Induced Repression: The Exception or the Rule in Developmental Signaling? *Developmental Cell*.

Agalioti, T., Lomvardas, S., Parekh, B., Yie, J., Maniatis, T., and Thanos, D., 2000. Ordered recruitment of chromatin modifying and general transcription factors to the IFN-beta promoter. *Cell* **103**(4), 667–678.

Ahmed, J.N., Ali, R.G., Warr, N.N., Wilson, H.M., Bellchambers, H.M., Barratt, K.S., Thompson, a J., and Arkell, R.M., 2013. A murine Zic3 transcript with a premature termination codon evades nonsense-mediated decay during axis formation. *Dis Model Mech* **6**(3), 755–767.

Akan, P., and Deloukas, P., 2008. DNA sequence and structural properties as predictors of human and mouse promoters. *Gene* **410**(1), 165–76.

Aker, M., Tubb, J., Groth, A.C., Bukovsky, A.A., Bell, A.C., Felsenfeld, G., Kiem, H.-P., Stamatoyannopoulos, G., and Emery, D.W., 2007. Extended core sequences from the cHS4 insulator are necessary for protecting retroviral vectors from silencing position effects. *Human gene therapy* **18**(4), 333–343.

Al-Souhibani, N., Al-Ahmadi, W., Hesketh, J.E., Blackshear, P.J., and Khabar, K.S.A., 2010. The RNA-binding zinc-finger protein tristetraprolin regulates AU-rich mRNAs involved in breast cancer-related processes. *Oncogene* **29**(29), 4205–4215.

Alten, L., Schuster-Gossler, K., Beckers, A., Groos, S., Ulmer, B., Hegermann, J., Ochs, M., and Gossler, A., 2012. Differential regulation of node formation, nodal ciliogenesis and cilia positioning by Noto and Foxj1. *Journal of Cell Science* **125**(7), e1–e1.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Research* **25**(17), 3389–3402.

Alvarez-Medina, R., Cayuso, J., Okubo, T., Takada, S., and Marti, E., 2007. Wnt canonical pathway restricts graded Shh/Gli patterning activity through the regulation of Gli3 expression. *Development* **135**(2), 237–247.

Andersen, C.L., Jensen, J.L., and Ørntoft, T.F., 2004. Normalization of Real-Time Quantitative Reverse Transcription-PCR Data: A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization, Applied to Bladder and Colon Cancer Data Sets. *Cancer Research* **64**(15), 5245–5250.

Anderson, R.M., Lawrence, A.R., Stottmann, R.W., Bachiller, D., and Klingensmith, J., 2002. Chordin and noggin promote organizing centers of forebrain development in the mouse. *Development* **129**(21), 4975–87.

Andersson, O., Reissmann, E., Jörnvall, H., and Ibáñez, C.F., 2006. Synergistic interaction between Gdf1 and Nodal during anterior axis development. *Developmental Biology* **293**(2), 370–381.

Andersson, R., 2014. Promoter or enhancer, what's the difference? Deconstruction of established distinctions and presentation of a unifying model. *BioEssays* **37**(3).

Andre, P., Song, H., Kim, W., Kispert, A., and Yang, Y., 2015. Wnt5a and Wnt11 regulate mammalian anterior-posterior axis elongation. *Development* **142**(8), 1516–27.

Ang, S.L., and Rossant, J., 1994. HNF-3 beta is essential for node and notochord formation in mouse development. *Cell* **78**(4), 561–74.

Ang, S.L., Wierda, A., Wong, D., Stevens, K.A., Cascio, S., Rossant, J., and Zaret, K.S., 1993. The formation and maintenance of the definitive endoderm lineage in the mouse: involvement of HNF3/forkhead proteins. *Development* **119**(4), 1301–15.

Antequera, F., and Bird, A., 1999. CpG islands as genomic footprints of promoters that are associated with replication origins. *Current Biology* **9**(17), R661–R667.

Aoto, K., Shikata, Y., Higashiyama, D., Shiota, K., and Motoyama, J., 2008. Fetal ethanol exposure activates protein kinase a and impairsShh expression in prechordal mesendoderm cells in the pathogenesis of holoprosencephaly. *Birth Defects Research Part A: Clinical and Molecular Teratology* **82**(4), 224–231.

Aoto, K., Shikata, Y., Imai, H., Matsumaru, D., Tokunaga, T., Shioda, S., Yamada, G., and Motoyama, J., 2009. Mouse Shh is required for prechordal plate maintenance during brain and craniofacial morphogenesis. *Developmental Biology* **327**(1), 106–120.

Arauz, R.F., Solomon, B.D., Pineda-Alvarez, D.E., Gropman, A.L., Parsons, J.A., Roessler, E., and Muenke, M., 2010. A hypomorphic allele in the FGF8 gene contributes to holoprosencephaly and is allelic to gonadotropin-releasing hormone deficiency in humans. *Molecular Syndromology* **1**(2), 59–66.

Arkell, R., and Beddington, R.S., 1997. BMP-7 influences pattern and growth of the developing hindbrain of mouse embryos. *Development* **124**(1).

Arkell, R.M., Fossat, N., and Tam, P.P.L., 2013. Wnt signalling in mouse gastrulation and anterior development: new players in the pathway and signal output. *Current opinion in genetics & development* **23**(4), 454–60.

Arkell, R.M., and Tam, P.P.L., 2012. Initiating head development in mouse embryos: integrating signalling and transcriptional activity. *Open Biology* **2**(3), 120030–120030.

Arnosti, D.N., and Kulkarni, M.M., 2005. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *Journal of Cellular Biochemistry* **94**(5), 890–898.

Aruga, J., Inoue, T., Hoshino, J., and Mikoshiba, K., 2002. Zic2 controls cerebellar development

in cooperation with Zic1. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **22**(1), 218–225.

Aruga, J., Kamiya, A., Takahashi, H., Fujimi, T.J., Shimizu, Y., Ohkawa, K., Yazawa, S., Umesono, Y., Noguchi, H., Shimizu, T., Saitou, N., Mikoshiba, K., Sakaki, Y., Agata, K., and Toyoda, A., 2006. A wide-range phylogenetic analysis of Zic proteins: Implications for correlations between protein structure conservation and body plan complexity. *Genomics* **87**(6), 783–792.

Au, K.S., Ashley-Koch, A., and Northrup, H., 2010. Epidemiologic and genetic aspects of spina bifida and other neural tube defects. *Developmental Disabilities Research Reviews* **16**(1), 6–15.

Bachiller, D., Klingensmith, J., Kemp, C., Belo, J.A., Anderson, R.M., May, S.R., McMahon, J.A., McMahon, A.P., Harland, R.M., Rossant, J., De Robertis, E.M., Bachiller, D., Klingensmith, J., Kemp, C., Belo, J.A., Anderson, R.M., May, S.R., McMahon, J.A., McMahon, A.P., Harland, R.M., and Rossant, J., 2000. The organizer factors Chordin and Noggin are required for mouse forebrain development. *Nature* **403**(6770), 658–661.

Bae, G.-U., Domené, S., Roessler, E., Schachter, K., Kang, J.-S., Muenke, M., and Krauss, R.S., 2011. Mutations in CDON, Encoding a Hedgehog Receptor, Result in Holoprosencephaly and Defective Interactions with Other Hedgehog Receptors. *The American Journal of Human Genetics* **89**(2), 231–240.

Baek, D., Villén, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P., 2008. The impact of microRNAs on protein output. *Nature* **455**(7209), 64–71.

Bakheet, T., Williams, B.R.G., and Khabar, K.S.A., 2006. ARED 3.0: the large and diverse AU-rich transcriptome. *Nucleic Acids Research* **34**(Database issue), D111–D114.

Banerji, J., Olson, L., and Schaffner, W., 1983. A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* **33**(3), 729–740.

Banerji, J., Rusconi, S., and Schaffner, W., 1981. Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**(2 Pt 1), 299–308.

Barkovich, A.J., and Quint, D.J., 1993. Middle interhemispheric fusion: An unusual variant of holoprosencephaly. *American Journal of Neuroradiology* **14**(2), 431–440.

Barolo, S., and Posakony, J.W., 2002. Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. *Genes & Development* **16**(10), 1167–81.

Barolo, S., Walker, R.G., Polyanovsky, a D., Freschi, G., Keil, T., and Posakony, J.W., 2000. A notch-independent activity of suppressor of hairless is required for normal mechanoreceptor physiology. *Cell* **103**(6), 957–969.

Barr, M., and Cohen, M.M., 2002. Autosomal recessive alobar holoprosencephaly with essentially normal faces. *American Journal of Medical Genetics* **112**(1), 28–30.

Barratt, K.S., Glanville-Jones, H.C., and Arkell, R.M., 2014. The Zic2 gene directs the formation and function of node cilia to control cardiac situs. *Genesis* **52**(6), 626–635.

Barreau, C., Paillard, L., and Osborne, H.B., 2005. AU-rich elements and associated factors: are there unifying principles? *Nucleic Acids Research* **33**(22), 7138–50.

Barrett, L.W., Fletcher, S., and Wilton, S.D., 2012. Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cellular and Molecular Life Sciences* **69**(21), 3613–3634.

Bay, S.N., and Caspary, T., 2012. What are those cilia doing in the neural tube? *Cilia* **1**(1), 19.

Beaudoing, E., Freier, S., Wyatt, J.R., Claverie, J.M., and Gautheret, D., 2000. Patterns of variant polyadenylation signal usage in human genes. *Genome Research* **10**(7), 1001–1010.

Beckers, A., Alten, L., Viebahn, C., Andre, P., and Gossler, A., 2007. The mouse homeobox gene Noto regulates node morphogenesis, notochordal ciliogenesis, and left right patterning. *Proceedings of the National Academy of Sciences of the United States of America* **104**(40), 15765–15770.

Bedard, J.E.J., Purnell, J.D., and Ware, S.M., 2007. Nuclear import and export signals are essential for proper cellular trafficking and function of ZIC3. *Human Molecular Genetics* **16**(2), 187–198.

Beddington, R.S., 1994. Induction of a second neural axis by the mouse node. *Development* **120**(3), 613–620.

Behringer, R.R., Bradley, A., Liu, P., Wakamiya, M., Shea, M.J., and Albrecht, U., 1999. Requirement for Wnt3 in vertebrate axis formation. *Nature Genetics* **22**(4), 361–365.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., and Haussler, D., 2004. Ultraconserved elements in the human genome. *Science* **304**(5675), 1321–1325.

Bellomo, D., Lander, A., Harragan, L., and Brown, N.A., 1996. Cell proliferation in mammalian gastrulation: The ventral node and notochord are relatively quiescent. *Developmental Dynamics* **205**(4), 471–485.

Belloni, E., Muenke, M., Roessler, E., Traverse, G., Siegel-Bartelt, J., Frumkin, A., Mitchell, H.F., Donis-Keller, H., Helms, C., Hing, A.V., Heng, H.H.Q., Koop, B., Martindale, D., Rommens, J.M., Tsui, L.-C., and Scherer, S.W., 1996. Identification of Sonic hedgehog as a candidate gene responsible for holoprosencephaly. *Nature Genetics* **14**(3), 353–356.

Belo, J.A., Bouwmeester, T., Leyns, L., Kertesz, N., Gallo, M., Follettie, M., and De Robertis, E.M., 1997. Cerberus-like is a secreted factor with neuralizing activity expressed in the anterior primitive endoderm of the mouse gastrula. *Mechanisms of Development* **68**(1–2), 45–57.

Belo, J.A., Leyns, L., Yamada, G., and De Robertis, E.M., 1998. The prechordal midline of the chondrocranium is defective in Goosecoid- 1 mouse mutants. *Mechanisms of Development* **72**(1–2), 15–25.

Ben-Haim, N., Lu, C., Guzman-Ayala, M., Pescatore, L., Mesnard, D., Bischofberger, M., Naef, F., Robertson, E.J., and Constam, D.B., 2006. The Nodal Precursor Acting via Activin Receptors Induces Mesoderm by Maintaining a Source of Its Convertases and BMP4. *Developmental Cell* **11**(3), 313–323.

Benko, S., Fantes, J.A., Amiel, J., Kleinjan, D.-J., Thomas, S., Ramsay, J., Jamshidi, N., Essafi, A., Heaney, S., Gordon, C.T., McBride, D., Golzio, C., Fisher, M., Perry, P., Abadie, V., Ayuso, C., Holder-Espinasse, M., Kilpatrick, N., Lees, M.M., Picard, A., Temple, I.K., Thomas, P., Vazquez, M.-P., Vekemans, M., Crollius, H.R., Hastie, N.D., Munnich, A., Etchevers, H.C., Pelet, A., Farlie, P.G., FitzPatrick, D.R., and Lyonnet, S., 2009. Highly conserved non-coding elements on either side of SOX9 associated with Pierre Robin sequence. *Nature Genetics* **41**(3), 359–364.

Berglund, A.C., Sjölund, E., Östlund, G., and Sonnhammer, E.L.L., 2008. InParanoid 6: Eukaryotic ortholog clusters with inparalogs. *Nucleic Acids Research* **36**(SUPPL. 1).

Bernard, D.J., Lee, K.B., and Santos, M.M., 2006. Activin B can signal through both ALK4 and ALK7 in gonadotrope cells. *Reproductive Biology and Endocrinology* **4**, 52.

Betel, D., Wilson, M., Gabow, A., Marks, D.S., and Sander, C., 2008. The microRNA.org resource: Targets and expression. *Nucleic Acids Research* **36**(SUPPL. 1).

Bjork, B.C., Fujiwara, Y., Davis, S.W., Qiu, H., Saunders, T.L., Sandy, P., Orkin, S., Camper, S.A., and Beier, D.R., 2010. A transient transgenic RNAi strategy for rapid characterization of gene function during embryonic development. *PLoS ONE* **5**(12), 1–11.

Blackshear, P.J., 2002. Tristetraprolin and other CCCH tandem zinc-finger proteins in the regulation of mRNA turnover. *Biochemical Society transactions* **30**(Pt 6), 945–52.

Blackwood, E.M., and Kadonaga, J.T., 1998. Going the distance: a current view of enhancer action. *Science* **281**(5373), 60–63.

Bogani, D., Willoughby, C., Davies, J., Kaur, K., Mirza, G., Paudyal, A., Haines, H., McKeone, R., Cadman, M., Pieles, G., Schneider, J.E., Bhattacharya, S., Hardy, A., Nolan, P.M., Tripodis, N., Depew, M.J., Chandrasekara, R., Duncan, G., Sharpe, P.T., Greenfield, A., Denny, P., Brown, S.D.M., Ragoussis, J., and Arkell, R.M., 2005. Dissecting the genetic complexity of human 6p deletion syndromes by using a region-specific, phenotype-driven mouse screen. *Proceedings of the National Academy of Sciences of the United States of America* **102**(35), 12477–12482.

Bonnafe, E., Touka, M., AitLounis, A., Baas, D., Barras, E., Ucla, C., Moreau, A., Flamant, F., Dubruille, R., Couble, P., Collignon, J., Durand, B., and Reith, W., 2004. The transcription factor RFX3 directs nodal cilium development and left-right asymmetry specification. *Molecular and Cellular Biology* **24**(10), 4417–4427.

Borok, M.J., Tran, D.A., Ho, M.C.W., and Drewell, R.A., 2010. Dissecting the regulatory switches of development: lessons from enhancer evolution in Drosophila. *Development* **137**(1), 5–13.

Borrmann, L., Wilkening, S., and Bullerdiek, J., 2001. The expression of HMGA genes is regulated by their 3'UTR. *Oncogene* **20**(33), 4537–4541.

Bradley, A., Evans, M., Kaufman, M.H., and Robertson, E., 1984. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**(5965), 255–256.

Brennan, C.M., and Steitz, J.A., 2001. HuR and mRNA stability. *Cellular and molecular life sciences : CMLS* **58**(2), 266–277.

Brennan, J., Norris, D.P., and Robertson, E.J., 2002. Nodal activity in the node governs left-right asymmetry. *Genes and Development* **16**(18), 2339–2344.

Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., and Cohen, S.M., 2003. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* **113**(1), 25–36.

Brennecke, J., Stark, A., Russell, R.B., and Cohen, S.M., 2005. Principles of microRNA-target recognition. *PLoS biology* **3**(3), e85.

Brewster, R., Lee, J., and Ruiz i Altaba, A., 1998. Gli/Zic factors pattern the neural plate by defining domains of cell differentiation. *Nature* **393**(6685), 579–83.

Britsch, S., Goerich, D.E., Riethmacher, D., Peirano, R.I., Rossner, M., Nave, K.A., Birchmeier, C., and Wegner, M., 2001. The transcription factor Sox10 is a key regulator of peripheral glial development. *Genes and Development* **15**(1), 66–78.

Brons, I.G.M., Smithers, L.E., Trotter, M.W.B., Rugg-Gunn, P., Sun, B., Chuva de Sousa Lopes, S.M., Howlett, S.K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R.A., and Vallier, L., 2007. Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**(7150), 191–195.

Brown, L.Y., Odent, S., David, V., Blayau, M., Dubourg, C., Apacik, C., Delgado, M.R.A., Hall, B.D., Reynolds, J.F., Sommer, A., Wieczorek, D., Brown, S.A., and Muenke, M., 2001. Holoprosencephaly due to mutations in ZIC2: alanine tract expansion mutations may be caused by parental somatic recombination. *Human Molecular Genetics* **10**(8), 791–796.

Brown, L.Y., Paraso, M., Arkell, R.M., and Brown, S., 2005. In vitro analysis of partial loss-of-function ZIC2 mutations in holoprosencephaly: Alanine tract expansion modulates DNA binding and transactivation. *Human Molecular Genetics* **14**(3), 411–420.

Brown, S., Warburton, D., Brown, L.Y., Yu, C.Y., Roeder, E., Stengel-Rutkowski, S., Hennekam, R.C.M., and Muenke, M., 1998. Holoprosencephaly due to mutations in ZIC2, a homologue of Drosophila odd-paired. *Nature Genetics* **20**(2), 180–183.

Brown, S.D.M., Nolan, P.M., Peters, J., Strivens, M., Rogers, D., Hagan, J., Spurr, N., Gray, I.C., Vizor, L., Brooker, D., Whitehill, E., Washbourne, R., Hough, T., Greenaway, S., Hewitt, M., Liu, X., McCormack, S., Pickford, K., Selley, R., Wells, C., Tymowska-Lalanne, Z., Roby, P., Glenister, P., Thornton, C., Thaung, C., Stevenson, J.-A., Arkell, R., Mburu, P., Hardisty, R., Kiernan, A., Erven, A., Steel, K.P., Voegeling, S., Guenet, J.-L., Nickols, C., Sadri, R., Naase, M., Isaacs, A., Davies, K., Browne, M., Fisher, E.M.C., Martin, J., Rastan, S., and Hunter, J., 2000. A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nature Genetics* **25**(4), 440–443.

Brueckner, M., 2007. Heterotaxia, congenital heart disease, and primary ciliary dyskinesia. *Circulation* **115**(22), 2793–5.

Bustin, S., Benes, V., and Garson, J., 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry* **55**(4), 611–22.

Cadigan, K.M., 2008. Wnt/β-Catenin Signaling: Turning the Switch. *Developmental Cell* **14**(3), 322–323.

Cadiñanos, J., and Bradley, A., 2007. Generation of an inducible and optimized piggyBac transposon systemy. *Nucleic Acids Research* **35**(12).

Calo, E., and Wysocka, J., 2013. Modification of Enhancer Chromatin: What, How, and Why? *Molecular Cell* **49**(5), 825–837.

Cambray, N., and Wilson, V., 2007. Two distinct sources for a population of maturing axial progenitors. *Development* **134**(15), 2829–2840.

Camus, A., Davidson, B.P., Billiards, S., Khoo, P., Rivera-Pérez, J.A., Wakamiya, M., Behringer, R.R., and Tam, P.P., 2000. The morphogenetic role of midline mesendoderm and ectoderm in the development of the forebrain and the midbrain of the mouse embryo. *Development* **127**(9), 1799–813.

Carey, M., 1998. The Enhanceosome and Transcriptional Synergy. *Cell* **92**(1), 5–8.

Carr, A., and Biggin, M.D., 1999. A comparison of invivo and invitro DNA-binding specificities

suggests a new model for homeoprotein DNA binding in Drosophila embryos. *The EMBO Journal* **18**(6), 1598–1608.

Carter, D., Chakalova, L., Osborne, C.S., Dai, Y., and Fraser, P., 2002. Long-range chromatin regulatory interactions in vivo. *Nature Genetics* **32**(4), 623–626.

Caspary, T., Larkins, C.E., and Anderson, K. V., 2007. The graded response to Sonic Hedgehog depends on cilia architecture. *Developmental cell* **12**(5), 767–78.

Chakraborty, S., Ji, H., Chen, J., Gersbach, C.A., and Leong, K.W., 2014. Vector modifications to eliminate transposase expression following piggyBac-mediated transgenesis. *Scientific Reports* **4**, 7403.

Chawengsaksophak, K., de Graaff, W., Rossant, J., Deschamps, J., and Beck, F., 2004. Cdx2 is essential for axial elongation in mouse development. *Proceedings of the National Academy of Sciences* **101**(20), 7641–7645.

Chawengsaksophak, K., James, R., Hammond, V.E., Köntgen, F., and Beck, F., 1997. Homeosis and intestinal tumours in Cdx2 mutant mice. *Nature* **386**(6620), 84–87.

Chen, C.-Y.A., and Shyu, A. Bin, 1995. AU-rich elements: characterization and importance in mRNA degradation. *Trends in Biochemical Sciences* **20**(11), 465–470.

Chen, C.-Y.A., Chen, T.M., and Shyu, A. Bin, 1994. Interplay of two functionally and structurally distinct domains of the c-fos AU-rich element specifies its mRNA-destabilizing function. *Molecular and Cellular Biology* **14**(1), 416–426.

Chen, C.Y., Gherzi, R., Ong, S.E., Chan, E.L., Raijmakers, R., Pruijn, G.J.M., Stoecklin, G., Moroni, C., Mann, M., and Karin, M., 2001. AU binding proteins recruit the exosome to degrade ARE-containing mRNAs. *Cell* **107**(4), 451–464.

Cheng, J.-H., Pan, D.Z.-C., Tsai, Z.T.-Y., and Tsai, H.-K., 2015. Genome-wide analysis of enhancer RNA in gene regulation across 12 mouse tissues. *Scientific Reports* **5**(1), 12648.

Cheng, X., Hsu, C., Currle, D.S., Hu, J.S., Barkovich, A.J., and Monuki, E.S., 2006. Central roles of the roof plate in telencephalic development and holoprosencephaly. *The Journal of Neuroscience* **26**(29), 7640–7649.

Chiang, C., Litingtung, Y., Lee, E., Young, K.E., Corden, J.L., Westphal, H., and Beachy, P. a, 1996. Cyclopia and defective axial patterning in mice lacking Sonic hedgehog gene function. *Nature* **383**(6599), 407–413.

Chin, A.J., Tsang, M., and Weinberg, E.S., 2000. Heart and gut chiralities are controlled independently from initial heart position in the developing zebrafish. *Developmental Biology* **227**, 403–421.

Choe, Y., Zarbalis, K.S., Pleasure, S.J., Iannarelli, P., and Dennehy, U., 2014. Neural Crest-Derived Mesenchymal Cells Require Wnt Signaling for Their Development and Drive Invagination of the Telencephalic Midline. *PLoS ONE* **9**(2), e86025.

Choksi, S.P., Lauter, G., Swoboda, P., and Roy, S., 2014. Switching on cilia: transcriptional networks regulating ciliogenesis. *Development* **141**(7), 1427–41.

Chu, G.C., Dunn, N.R., Anderson, D.C., Oxburgh, L., and Robertson, E.J., 2004. Differential requirements for Smad4 in TGFbeta-dependent patterning of the early mouse embryo. *Development* **131**(15), 3501–3512.

Cohen, M.M., 2012. Perspectives on asymmetry: The Erickson lecture. *American Journal of Medical Genetics, Part A* **158 A**(12), 2981–2998.

Collier, L.S., Carlson, C.M., Ravimohan, S., Dupuy, A.J., and Largaespada, D. a, 2005. Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature* **436**(7048), 272–276.

Collignon, J., Varlet, I., and Robertson, E.J., 1996. Relationship between asymmetric Nodal expression and the direction of embryonic turning. *Nature* **381**(6578), 155–158.

Conlon, F.L., Lyons, K.M., Takaesu, N., Barth, K.S., Kispert, A., Herrmann, B., and Robertson, E.J., 1994. A primary requirement for Nodal in the formation and maintenance of the primitive streak in the mouse. *Development* **120**(7), 1919–28.

Costello, I., Nowotschin, S., Sun, X., Mould, A.W., Hadjantonakis, A.-K., Bikoff, E.K., and Robertson, E.J., 2015. Lhx1 functions together with Otx2, Foxa2, and Ldb1 to govern anterior mesendoderm, node, and midline development. *Genes & Development* **29**(20), 2108–22.

Courey, A., and Jia, S., 2001. Transcriptional repression: the long and the short of it. *Genes & Development*.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., Boyer, L.A., Young, R.A., and Jaenisch, R., 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America* **107**(50), 21931–21936.

Crossley, P.H., and Martin, G.R., 1995. The mouse Fgf8 gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo. *Development* **121**(2), 439–451.

Cruz, C., Ribes, V., Kutejova, E., Cayuso, J., Lawson, V., Norris, D.P., Stevens, J.L., Davey, M., Blight, K., Bangs, F., Mynett, A., Hirst, E., Chung, R., Balaskas, N., Brody, S.L., Marti, E., and Briscoe, J., 2010. Foxj1 regulates floor plate cilia architecture and modifies the response of cells to Sonic Hedgehog signalling. *Development* **137**(24), 4271–4282.

Currle, D.S., Cheng, X., Hsu, C.-M., and Monuki, E.S., 2005. Direct and indirect roles of CNS dorsal midline cells in choroid plexus epithelia formation. *Development* **132**(15), 3549–3559.

Dailey, L., Ambrosetti, D., Mansukhani, A., and Basilico, C., 2005. Mechanisms underlying differential responses to FGF signaling. *Cytokine & Growth Factor Reviews* **16**(2), 233–247.

Dale, J.K., Vesque, C., Lints, T.J., Sampath, T.K., Furley, A., Dodd, J., and Placzek, M., 1997. Cooperation of BMP7 and SHH in the induction of forebrain ventral midline cells by prechordal mesoderm. *Cell* **90**(2), 257–69.

Danckwardt, S., Hentze, M.W., and Kulozik, A.E., 2008. 3' end mRNA processing: molecular mechanisms and implications for health and disease. *The EMBO journal* **27**(3), 482–498.

Darrow, E.M., and Chadwick, B.P., 2013. Boosting transcription by transcription: Enhancer-associated transcripts. *Chromosome Research* **21**(6–7), 713–724.

Davidson, B.P., and Tam, P.P.L., 2000. The node of the mouse embryo. *Current Biology* **10**(17).

Davis, A., Amin, N.M., Johnson, C., Bagley, K., Ghashghaei, H.T., and Nascone-Yoder, N., 2017. Stomach curvature is generated by left-right asymmetric gut morphogenesis. *Development* **144**(8).

De la Cruz, J.M., Bamford, R.N., Burdine, R.D., Roessler, E., Barkovich, J.A., Donnai, D., Schier, A.F., and Muenke, M., 2002. A loss-of-function mutation in the CFC domain of TDGF1 is associated with human forebrain defects. *Human Genetics* **110**(5), 422–428.

DeMyer, W., Zeman, W., and Palmer, C.G., 1964. The face predicts the brain: Diagnostic significance of median facial anomalies for holoprosencephaly (arhinencephaly). *Pediatrics* **34**(August), 256–263.

Dennler, S., Itoh, S., Vivien, D., Dijke, P. Ten, Huet, S., and Gauthier, J.M., 1998. Direct binding of Smad3 and Smad4 to critical TGFβ-inducible elements in the promoter of human plasminogen activator inhibitor-type 1 gene. *EMBO Journal* **17**(11), 3091–3100.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M., and Gascuel, O., 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research* **36**(Web Server issue).

Deveci, M., Catalyürek, U. V, and Toland, A.E., 2014. mrSNP: software to detect SNP effects on microRNA binding. *BMC Bioinformatics* **15**, 73.

Didon, L., Zwick, R.K., Chao, I.W., Walters, M.S., Wang, R., Hackett, N.R., and Crystal, R.G., 2013. RFX3 modulation of FOXJ1 regulation of cilia genes in the human airway epithelium. *Respiratory Research* **14**(1), 70.

Ding, S., Wu, X., Li, G., Han, M., Zhuang, Y., and Xu, T., 2005. Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* **122**(3), 473–483.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G.K., Khatun, J., Williams, B.A., Zaleski, C., Rozowsky, J., Röder, M., Kokocinski, F., Abdelhamid, R.F., Alioto, T., Antoshechkin, I., Baer, M.T., Bar, N.S., Batut, P., Bell, K., Bell, I., Chakrabortty, S., Chen, X., Chrast, J., Curado, J., Derrien, T., Drenkow, J., Dumais, E., Dumais, J., Duttagupta, R., Falconnet, E., Fastuca, M., Fejes-Toth, K., Ferreira, P., Foissac, S., Fullwood, M.J., Gao, H., Gonzalez, D., Gordon, A., Gunawardena, H., Howald, C., Jha, S., Johnson, R., Kapranov, P., King, B., Kingswood, C., Luo, O.J., Park, E., Persaud, K., Preall, J.B., Ribeca, P., Risk, B., Robyr, D., Sammeth, M., Schaffer, L., See, L.-H., Shahab, A., Skancke, J., Suzuki, A.M., Takahashi, H., Tilgner, H., Trout, D., Walters, N., Wang, H., Wrobel, J., Yu, Y., Ruan, X., Hayashizaki, Y., Harrow, J., Gerstein, M., Hubbard, T., Reymond, A., Antonarakis, S.E., Hannon, G.J., Giddings, M.C., Ruan, Y., Wold, B., Carninci, P., Guigó, R., and Gingeras, T.R., 2012. Landscape of transcription in human cells. *Nature* **489**(7414), 101–8.

Doetschman, T., 2009. Influence of genetic background on genetically engineered mouse phenotypes. *Methods in molecular biology (Clifton, N.J.)* **530**, 423–33.

Doherty, J.E., Huye, L.E., Yusa, K., Zhou, L., Craig, N.L., and Wilson, M.H., 2012. Hyperactive piggyBac gene transfer in human cells and in vivo. *Human gene therapy* **23**(3), 311–20.

Dou, C., Lee, J., Liu, B., Liu, F., Massague, J., Xuan, S., and Lai, E., 2000. BF-1 interferes with transforming growth factor beta signaling by associating with Smad partners. *Molecular and Cellular Biology* **20**(17), 6201–11.

Dou, C.L., Li, S., and Lai, E., 1999. Dual role of brain factor-1 in regulating growth and patterning of the cerebral hemispheres. *Cerebral Cortex* **9**(6), 543–50.

Dubourg, C., Bendavid, C., Pasquier, L., Henry, C., Odent, S., and David, V., 2007. Holoprosencephaly. *Orphanet Journal of Rare Diseases* **2**, 8.

Dubourg, C., Carré, W., Hamdi-Rozé, H., Mouden, C., Roume, J., Abdelmajid, B., Amram, D., Baumann, C., Chassaing, N., Coubes, C., Faivre-Olivier, L., Ginglinger, E., Gonzales, M., Levy-Mozziconacci, A., Lynch, S.A., Naudion, S., Pasquier, L., Poidvin, A., Prieur, F., Sarda, P., Toutain, A., Dupé, V., Akloul, L., Odent, S., de Tayrac, M., and David, V., 2016. Mutational Spectrum in Holoprosencephaly Shows That FGF is a New Major Signaling Pathway. *Human Mutation* **37**(12), 1329–1339.

Dubourg, C., David, V., Gropman, A., Mercier, S., Muenke, M., Odent, S., Pineda-Alvarez, D.E., and Roessler, E., 2011. Clinical utility gene card for: Holoprosencephaly. *European Journal of Human Genetics* **19**(1), 118–20.

DuBridge, R.B., Tang, P., Hsia, H.C., Leong, P.M., Miller, J.H., and Calos, M.P., 1987. Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system. *Molecular and cellular biology* **7**(1), 379–87.

Dudley, A.T., Lyons, K.M., and Robertson, E.J., 1995. A requirement for bone morphogenetic protein-7 during development of the mammalian kidney and eye. *Genes & Development* **9**(22), 2795–807.

Dufort, D., Schwartz, L., Harpal, K., and Rossant, J., 1998. The transcription factor HNF3beta is required in visceral endoderm for normal primitive streak morphogenesis. *Development* **125**(16), 3015–3025.

Dunn, N.R., Vincent, S.S.D., Oxburgh, L., Robertson, E.E.J., and Bikoff, E.K.E., 2004. Combinatorial activities of Smad2 and Smad3 regulate mesoderm formation and patterning in the mouse embryo. *Development* **131**(8), 1717–1728.

Dupé, V., Rochard, L., Mercier, S., Le Pétillon, Y., Gicquel, I., Bendavid, C., Bourrouillou, G., Kini, U., Thauvin-Robinet, C., Bohan, T.P., Odent, S., Dubourg, C., and David, V., 2011. NOTCH, a new signaling pathway implicated in holoprosencephaly. *Human Molecular Genetics* **20**(6), 1122–31.

Dupuy, A.J., Clark, K., Carlson, C.M., Fritz, S., Davidson, A.E., Markley, K.M., Finley, K., Fletcher, C.F., Ekker, S.C., Hackett, P.B., Horn, S., and Largaespada, D. a, 2002. Mammalian germ-line transgenesis by transposition. *Proceedings of the National Academy of Sciences of the United States of America* **99**(7), 4495–9.

Echelard, Y., Epstein, D.J., St-Jacques, B., Shen, L., Mohler, J., McMahon, J.A., and McMahon, A.P., 1993. Sonic Hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of CNS polarity. *Cell* **75**(7), 1417–30.

Elms, P., Scurry, A., Davies, J., Willoughby, C., Hacker, T., Bogani, D., and Arkell, R.M., 2004. Overlapping and distinct expression domains of Zic2 and Zic3 during mouse gastrulation. *Gene Expression Patterns* **4**(5), 505–511.

Elms, P., Siggers, P., Napper, D., Greenfield, A., and Arkell, R.M., 2003. Zic2 is required for neural crest formation and hindbrain patterning during mouse development. *Developmental Biology* **264**(2), 391–406.

Episkopou, V., Arkell, R., Timmons, P.M., Walsh, J.J., Andrew, R.L., and Swan, D., 2001. Induction of the mammalian node requires Arkadia function in the extraembryonic lineages. *Nature* **410**(6830), 825–830.

Ermakov, A., Stevens, J.L., Whitehill, E., Robson, J.E., Pieles, G., Brooker, D., Goggolidou, P.,

Powles-Glover, N., Hacker, T., Young, S.R., Dear, N., Hirst, E., Tymowska-Lalanne, Z., Briscoe, J., Bhattacharya, S., and Norris, D.P., 2009. Mouse mutagenesis identifies novel roles for left-right patterning genes in pulmonary, craniofacial, ocular, and limb development. *Developmental Dynamics* **238**(3), 581–594.

Evans, M.J., and Kaufman, M.H., 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**(5819), 154–156.

Fabian, M.R., Sonenberg, N., and Filipowicz, W., 2010. Regulation of mRNA translation and stability by microRNAs. *Annual review of biochemistry* **79**, 351–379.

Fan, X. C., and Steitz, J. A., 1998a. Overexpression of HuR, a nuclear-cytoplasmic shuttling protein, increases the in vivo stability of ARE-containing mRNAs. *The EMBO journal* **17**(12), 3448–60.

Fan, X. C., and Steitz, J. A., 1998b. HNS, a nuclear-cytoplasmic shuttling sequence in HuR. *Proceedings of the National Academy of Sciences* **95**(26), 15293–15298.

Farley, E.K., Olson, K.M., Zhang, W., Rokhsar, D.S., and Levine, M.S., 2016. Syntax compensates for poor binding sites to encode tissue specificity of developmental enhancers. *Proceedings of the National Academy of Sciences of the United States of America* **113**(23), 6508–13.

Fernandes, M., Gutin, G., Alcorn, H., McConnell, S.K., and Hebert, J.M., 2007. Mutations in the BMP pathway in mice support the existence of two molecular classes of holoprosencephaly. *Development* **134**(21), 3789–3794.

Ferrer-Vaquer, A., Piliszek, A., Tian, G., Aho, R.J., Dufort, D., and Hadjantonakis, A.-K., 2010. A sensitive and bright single-cell resolution live imaging reporter of Wnt/ß-catenin signaling in the mouse. *BMC Developmental Biology* **10**(1), 121.

Field, S., Riley, K.L., Grimes, D.T., Hilton, H., Simon, M., Powles-Glover, N., Siggers, P., Bogani, D., Greenfield, A., and Norris, D.P., 2011. Pkd1l1 establishes left-right asymmetry and physically interacts with Pkd2. *Development* **138**(6), 1131–1142.

Filosa, S., Rivera-Pérez, J.A., Gómez, A.P., Gansmuller, A., Sasaki, H., Behringer, R.R.R., Ang, S.L.L., Rivera-Perez, J.A., Gomez, A.P., Gansmuller, A., Sasaki, H., Behringer, R.R.R., and Ang, S.L.L., 1997. Goosecoid and HNF-3beta genetically interact to regulate neural tube patterning during mouse embryogenesis. *Development* **124**(14), 2843–54.

Finger, J.H., Smith, C.M., Hayamizu, T.F., McCright, I.J., Eppig, J.T., Kadin, J.A., Richardson, J.E., and Ringwald, M., 2011. The mouse Gene Expression Database (GXD): 2011 update. *Nucleic Acids Research* **39**(SUPPL. 1), D835-41.

Fraser, M.J., Ciszczon, T., Elick, T., and Bauser, C., 1996. Precise excision of TTAA-specific lepidopteran transposons piggyBac (IFP2) and tagalong (TFP3) from the baculovirus genome in cell lines from two species of Lepidoptera. *Insect molecular biology* **5**(2), 141–51.

Frazer, K.A., Elnitski, L., Church, D.M., Dubchak, I., and Hardison, R.C., 2003. Cross-species sequence comparisons: a review of methods and available resources. *Genome research* **13**(1), 1–12.

Fu, Y., Foden, J.A., Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K., and Sander, J.D., 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nature Biotechnology* **31**(9), 822–6.

Fuccillo, M., Rallu, M., McMahon, A.P., and Fishell, G., 2004. Temporal requirement for hedgehog signaling in ventral telencephalic patterning. *Development* **131**(20).

Fujimi, T.J., Hatayama, M., and Aruga, J., 2012. Xenopus Zic3 controls notochord and organizer development through suppression of the Wnt/Beta-catenin signaling pathway. *Developmental Biology* **361**(2), 220–231.

Furushima, K., Murata, T., Matsuo, I., and Aizawa, S., 2000. A new murine zinc finger gene, Opr. *Mechanisms of Development* **98**(1–2), 161–164.

Furuta, Y., Piston, D.W., and Hogan, B.L., 1997. Bone morphogenetic proteins (BMPs) as regulators of dorsal forebrain development. *Development* **124**(11), 2203–12.

Gao, N., LeLay, J., Vatamaniuk, M.Z., Rieck, S., Friedman, J.R., and Kaestner, K.H., 2008. Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. *Genes & Development* **22**(24), 3435–48.

Garneau, N.L., Wilusz, J., and Wilusz, C.J., 2007. The highways and byways of mRNA decay. *Nature reviews. Molecular cell biology* **8**(2), 113–126.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V., 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **109**(39), E2579-86.

Geanacopoulos, M., Vasmatzis, G., Lewis, D.E.A., Roy, S., Lee, B., and Adhya, S., 1999. GalR mutants defective in repressosome formation. *Genes & Development* **13**(10), 1251–1262.

Geisberg, J. V., Moqtaderi, Z., Fan, X., Ozsolak, F., and Struhl, K., 2014. Global analysis of mRNA isoform half-lives reveals stabilizing and destabilizing elements in yeast. *Cell* **156**(4), 812–824.

Gelb, B.D., 2004. Genetic basis of congenital heart disease. *Current Opinion in Cardiology* **19**(2), 110–115.

Giraldez, A.J., Cinalli, R.M., Glasner, M.E., Enright, A.J., Thomson, J.M., Baskerville, S., Hammond, S.M., Bartel, D.P., and Schier, A.F., 2005. MicroRNAs regulate brain morphogenesis in zebrafish. *Science* **308**(5723), 833–838.

Giraldo, P., and Montoliu, L., 2001. Size matters: use of YACs, BACs and PACs in transgenic animals. *Transgenic research* **10**(2), 83–103.

Gong, J., Liu, C., Liu, W., Wu, Y., Ma, Z., Chen, H., and Guo, A.-Y., 2015. An update of miRNASNP database for better SNP selection by GWAS data, miRNA expression and online tools. *Database* **2015**.

Gong, S., Zheng, C., Doughty, M.L., Losos, K., Didkovsky, N., Schambra, U.B., Nowak, N.J., Joyner, A., Leblanc, G., Hatten, M.E., and Heintz, N., 2003. A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* **425**(6961), 917–925.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R., 2010. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic Acids Research* **38**(SUPPL. 2).

Goulding, M.D., Chalepakis, G., Deutsch, U., Erselius, J.R., and Gruss, P., 1991. Pax-3, a novel murine DNA binding protein expressed during early neurogenesis. *The EMBO journal* **10**(5), 1135–47.

Gowri, P.M., Yu, J.H., Shaufl, A., Sperling, M.A., and Menon, R.K., 2003. Recruitment of a repressosome complex at the growth hormone receptor promoter and its potential role in diabetic nephropathy. *Molecular and Cellular Biology* **23**(3), 815–25.

GraphPad Software, I., 2017. GraphPad [WWW Document]. URL www.graphpad.com (accessed 1.1.17).

Greene, N.D.E., and Copp, A.J., 2009. Development of the vertebrate central nervous system: formation of the neural tube. *Prenatal diagnosis* **29**(4), 303–11.

Gripp, K.W., Wotton, D., Edwards, M.C., Roessler, E., Ades, L., Meinecke, P., Richieri-Costa, A., Zackai, E.H., Massagué, J., Muenke, M., and Elledge, S.J., 2000. Mutations in TGIF cause holoprosencephaly and link NODAL signalling to human neural axis determination. *Nature Genetics* **25**(2), 205–8.

Gross, D.S., and Garrard, W.T., 1988. Nuclease hypersensitive sites in chromatin. *Annual review of biochemistry* **57**, 159–197.

Groth, A.C., Liu, M., Wang, H., Lovelett, E., and Emery, D.W., 2013. Identification and Characterization of Enhancer-Blocking Insulators to Reduce Retroviral Vector Genotoxicity. *PLoS ONE* **8**(10), e76528.

Groves, A.K., and LaBonne, C., 2014. Setting appropriate boundaries: Fate, patterning and competence at the neural plate border. *Developmental Biology* **389**(1), 2–12.

Gruber, A.R., Fallmann, J., Kratochvill, F., Kovarik, P., and Hofacker, I.L., 2011. AREsite: A database for the comprehensive investigation of AU-rich elements. *Nucleic Acids Research* **39**(SUPPL. 1).

Gualdi, R., Bossard, P., Zheng, M., Hamada, Y., Coleman, J.R., and Zaret, K.S., 1996. Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes & Development* **10**(13), 1670–82.

Guo, G., Yang, J., Nichols, J., Hall, J.S., Eyres, I., Mansfield, W., and Smith, A., 2009. Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* **136**(7), 1063–1069.

Gupta, S., and Sen, J., 2016. Roof plate mediated morphogenesis of the forebrain: New players join the game. *Developmental Biology* **413**(2), 145–152.

Gutin, G., Fernandes, M., Palazzolo, L., Paek, H., Yu, K., Ornitz, D.M., McConnell, S.K., and Hébert, J.M., 2006. FGF signalling generates ventral telencephalic cells independently of SHH. *Development* **133**(15), 2937–2946.

Hadjantonakis, A.-K., Pisano, E., and Papaioannou, V.E., 2008. Tbx6 regulates left/right patterning in mouse embryos through effects on nodal cilia and perinodal signaling. *PLoS ONE* **3**(6), e2511.

Hagiwara, H., Ohwada, N., and Takata, K., 2004. Cell biology of normal and abnormal ciliogenesis in the ciliated epithelium. *International Review of Cytology* **234**, 101–41.

Halder, G., Polaczyk, P., Kraus, M.E., Hudson, A., Kim, J., Laughon, A., and Carroll, S., 1998. The Vestigial and Scalloped proteins act together to directly regulate wing-specific gene expression in Drosophila. *Genes & Development* **12**(24), 3900–9.

Halees, A.S., El-badrawi, R., and Khabar, K.S.A., 2008. ARED Organism: Expansion of ARED reveals AU-rich element cluster variations between human and mouse. *Nucleic Acids*

Research **36**(SUPPL. 1).

Hallonet, M., Kaestner, K.H., Martin-Parras, L., Sasaki, H., Betz, U.A.K., and Ang, S.-L., 2002. Maintenance of the Specification of the Anterior Definitive Endoderm and Forebrain Depends on the Axial Mesendoderm: A Study Using HNF3β/Foxa2 Conditional Mutants. *Developmental Biology* **243**(1), 20–33.

Hanashima, C., Li, S.C., Shen, L., Lai, E., and Fishell, G., 2004. Foxg1 suppresses early cortical cell fate. *Science* **303**(5654), 56–59.

Harteveld, C.L., Losekoot, M., Haak, H., Heister, G.A., Giordano, P.C., and Bernini, L.F., 1994. A novel polyadenylation signal mutation in the alpha 2-globin gene causing alpha thalassaemia. *Br J Haematol* **87**(1), 139–143.

Harvey, R.P., 2002. Patterning the vertebrate heart. *Nature Reviews Genetics* **3**(7), 544–556.

Hatayama, M., Tomizawa, T., Sakai-Kato, K., Bouvagnet, P., Kose, S., Imamoto, N., Yokoyama, S., Utsunomiya-Tate, N., Mikoshiba, K., Kigawa, T., and Aruga, J., 2008. Functional and structural basis of the nuclear localization signal in the ZIC3 zinc finger domain. *Human Molecular Genetics* **17**(22), 3459–3473.

Haycraft, C.J., Banizs, B., Aydin-Son, Y., Zhang, Q., Michaud, E.J., and Yoder, B.K., 2005. Gli2 and Gli3 localize to cilia and require the intraflagellar transport protein polaris for processing and function. *PLoS Genetics* **1**(4).

Hayhurst, M., Gore, B.B., Tessier-Lavigne, M., and McConnell, S.K., 2008. Ongoing sonic hedgehog signaling is required for dorsal midline formation in the developing forebrain. *Developmental Neurobiology* **68**(1), 83–100.

Hébert, J.M., and Fishell, G., 2008. The genetics of early telencephalon patterning: some assembly required. *Nature reviews. Neuroscience* **9**(9), 678–85.

Hébert, J.M., Hayhurst, M., Marks, M.E., Kulessa, H., Hogan, B.L.M., and McConnell, S.K., 2003. BMP ligands act redundantly to pattern the dorsal telencephalic midline. *Genesis* **35**(4), 214–219.

Hébert, J.M., Mishina, Y., and McConnell, S.K., 2002. BMP signaling is required locally to pattern the dorsal telencephalic midline. *Neuron* **35**(6).

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., Ching, K.A., Antosiewicz-Bourget, J.E., Liu, H., Zhang, X., Green, R.D., Lobanenkov, V. V, Stewart, R., Thomson, J.A., Crawford, G.E., Kellis, M., and Ren, B., 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**(7243), 108–112.

Heintzman, N.D., Stuart, R.K., Hon, G.C., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., Wang, W., Weng, Z., Green, R.D., Crawford, G.E., and Ren, B., 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nature Genetics* **39**(3), 311–318.

Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., and Vandesompele, J., 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome biology* **8**(2), R19.

Helwak, A., Kudla, G., Dudnakova, T., and Tollervey, D., 2013. Mapping the human miRNA interactome by CLASH reveals frequent noncanonical binding. *Cell* **153**(3), 654–65.

Herbst, R.S., Friedman, N., Darnell, J.E., and Babiss, L.E., 1989. Positive and negative regulatory elements in the mouse albumin enhancer. *Proceedings of the National Academy of Sciences of the United States of America* **86**(5), 1553–7.

Higgs, D.R., Goodbourn, S.E.Y., Lamb, J., Clegg, J.B., Weatherall, D.J., and Proudfoot, N.J., 1983. α-Thalassaemia caused by a polyadenylation signal mutation. *Nature* **306**(5941), 398–400.

Himeda, C.L., Barro, M. V., and Emerson, C.P., 2013. Pax3 synergizes with Gli2 and Zic1 in transactivating the Myf5 epaxial somite enhancer. *Developmental Biology* **383**(1), 7–14.

Hiraiwa, N., Ishimoto, M., and Yasue, H., 2013. Examination of the mouse embryo by micro-CT. *Experimental Animals* **62**(1), 57–61.

Hirokawa, N., Tanaka, Y., Okada, Y., and Takeda, S., 2006. Nodal flow and the generation of left-right asymmetry. *Cell* **125**(1), 33–45.

Hitti, E., 2012. Sequence variations affecting AU-rich element function and disease. *Frontiers in Bioscience* **17**(1), 1846.

Hoch, R. V., Rubenstein, J.L.R., and Pleasure, S., 2009. Genes and signaling events that establish regional patterning of the mammalian forebrain. *Seminars in Cell and Developmental Biology* **20**(4), 378–386.

Hochschild, A., and Dove, S.L., 1998. Protein–Protein Contacts that Activate and Repress Prokaryotic Transcription. *Cell* **92**(5), 597–600.

Hong, M., Krauss, R.S., Hannenhalli, S., Wang, Z., and Cappola, T., 2012. Cdon Mutation and Fetal Ethanol Exposure Synergize to Produce Midline Signaling Defects and Holoprosencephaly Spectrum Disorders in Mice. *PLoS Genetics* **8**(10), e1002999.

Hoodless, P.A., Pye, M., Chazaud, C., Labbé, E., Attisano, L., Rossant, J., and Wrana, J.L., 2001. FoxH1 (Fast) functions to specify the anterior primitive streak in the mouse. *Genes & Development* **15**(10), 1257–71.

Horie, K., Yusa, K., Yae, K., Odajima, J., Fischer, S.E.J., Keng, V.W., Hayakawa, T., Mizuno, S., Kondoh, G., Ijiri, T., Matsuda, Y., Plasterk, R.H. a, and Takeda, J., 2003. Characterization of Sleeping Beauty transposition and its application to genetic screening in mice. *Molecular and Cellular Biology* **23**(24), 9189–207.

Houtmeyers, R., Souopgui, J., Tejpar, S., and Arkell, R.M., 2013. The ZIC gene family encodes multi-functional proteins essential for patterning and morphogenesis. *Cellular and Molecular Life Sciences* **70**(20), 3791–3811.

Houtmeyers, R., Tchouate Gainkam, O., Glanville-Jones, H.C., Van den Bosch, B., Chappell, A., Barratt, K.S., Souopgui, J., Tejpar, S., and Arkell, R.M., 2016. Zic2 mutation causes Holoprosencephaly via disruption of NODAL signalling. *Human Molecular Genetics* **61**.

Hruz, T., Wyss, M., Docquier, M., Pfaffl, M.W., Masanetz, S., Borghi, L., Verbrugghe, P., Kalaydjieva, L., Bleuler, S., Laule, O., Descombes, P., Gruissem, W., and Zimmermann, P., 2011. RefGenes: identification of reliable and condition specific reference genes for RT-qPCR data normalization. *BMC Genomics* **12**, 156.

Huang, X., Guo, H., Tammana, S., Jung, Y.-C., Mellgren, E., Bassi, P., Cao, Q., Tu, Z.J., Kim, Y.C., Ekker, S.C., Wu, X., Wang, S.M., and Zhou, X., 2010. Gene transfer efficiency and genome-wide integration profiling of Sleeping Beauty, Tol2, and piggyBac transposons in human primary T cells. *Molecular therapy : the journal of the American Society of Gene Therapy*

18(10), 1803–13.

Huangfu, D., and Anderson, K. V., 2005. Cilia and Hedgehog responsiveness in the mouse. *Proceedings of the National Academy of Sciences of the United States of America* **102**(32), 11325–30.

Hughes, J.R., Roberts, N., McGowan, S., Hay, D., Giannoulatou, E., Lynch, M., De Gobbi, M., Taylor, S., Gibbons, R., and Higgs, D.R., 2014. Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nature Genetics* **46**(2), 205–12.

Ibañez-Tallon, I., Pagenstecher, A., Fliegauf, M., Olbrich, H., Kispert, A., Ketelsen, U.-P., North, A., Heintz, N., and Omran, H., 2004. Dysfunction of axonemal dynein heavy chain Mdnah5 inhibits ependymal flow and reveals a novel mechanism for hydrocephalus formation. *Human Molecular Genetics* **13**(18), 2133–41.

Inoue, T., Hatayama, M., Tohmonda, T., Itohara, S., Aruga, J., and Mikoshiba, K., 2004. Mouse Zic5 deficiency results in neural tube defects and hypoplasia of cephalic neural crest derivatives. *Developmental Biology* **270**(1), 146–162.

Ip, C.K., Fossat, N., Jones, V., Lamonerie, T., and Tam, P.P.L., 2014. Head formation: OTX2 regulates Dkk1 and Lhx1 activity in the anterior mesendoderm. *Development* **141**(20), 3859–3867.

Ivics, Z., Hackett, P.B., Plasterk, R.H., and Izsvák, Z., 1997. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* **91**(4), 501–10.

Jacobi, A.M., Rettig, G.R., Turk, R., Collingwood, M.A., Zeiner, S.A., Quadros, R.M., Harms, D.W., Bonthuis, P.J., Gregg, C., Ohtsuka, M., Gurumurthy, C.B., and Behlke, M.A., 2017. Simplified CRISPR tools for efficient genome editing and streamlined protocols for their delivery into mammalian cells and mouse zygotes. *Methods* **121**–**122**, 16–28.

Jankovic, L., Efremov, G.D., Petkov, G., Kattamis, C., George, E., Yang, K.G., Stoming, T.A., and Huisman, T.H., 1990. Two novel polyadenylation mutations leading to beta(+)-thalassemia. *Br J Haematol* **75**(1), 122–126.

Jeong, Y., and Epstein, D.J., 2003. Distinct regulators of Shh transcription in the floor plate and notochord indicate separate origins for these tissues in the mouse node. *Development* **130**(16), 3891–3902.

Jeong, Y., Leskow, F.C., El-Jaick, K.B.K., Roessler, E., Muenke, M., Yocum, A., Dubourg, C.C., Li, X., Geng, X., Oliver, G., and Epstein, D.J., 2008. Regulation of a remote Shh forebrain enhancer by the Six3 homeoprotein. *Nature Genetics* **40**(11), 1348–53.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E., 2012. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science*.

Jing, Q., Huang, S., Guth, S., Zarubin, T., Motoyama, A., Chen, J., Di Padova, F., Lin, S.C., Gram, H., and Han, J., 2005. Involvement of MicroRNA in AU-Rich Element-Mediated mRNA Instability. *Cell* **120**(5), 623–634.

John, B., Enright, A.J., Aravin, A., Tuschl, T., Sander, C., and Marks, D.S., 2004. Human microRNA targets. *PLoS Biology* **2**(11).

Kaikkonen, M.U., Lam, M.T.Y., and Glass, C.K., 2011. Non-coding RNAs as regulators of gene

expression and epigenetics. *Cardiovascular Research* **90**(3), 430–440.

Kamura, K., Kobayashi, D., Uehara, Y., Koshida, S., Iijima, N., Kudo, A., Yokoyama, T., and Takeda, H., 2011. Pkd1l1 complexes with Pkd2 on motile cilia and functions to establish the left-right axis. *Development* **138**(6), 1121–1129.

Kang, H.S., Beak, J.Y., Kim, Y.-S., Herbert, R., and Jetten, A.M., 2009. Glis3 is associated with primary cilia and Wwtr1/TAZ and implicated in polycystic kidney disease. *Molecular and Cellular Biology* **29**(10), 2556–69.

Kang, H.S., ZeRuth, G., Lichti-Kaiser, K., Vasanth, S., Yin, Z., Kim, Y.-S., and Jetten, A.M., 2010. Gli-similar (Glis) Krüppel-like zinc finger proteins: insights into their physiological functions and critical roles in neonatal diabetes and cystic renal disease. *Histology and Histopathology* **25**(11), 1481–96.

Katzman, S., Kern, A.D., Bejerano, G., Fewell, G., Fulton, L., Wilson, R.K., Salama, S.R., and Haussler, D., 2007. Human genome ultraconserved elements are ultraselected. *Science* **317**(5840), 915.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Meintjes, P., and Drummond, A., 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**(12), 1647–1649.

Kemp, C., Willems, E., Abdo, S., Lambiv, L., and Leyns, L., 2005. Expression of all Wnt genes and their secreted antagonists during mouse blastocyst and postimplantation development. *Developmental Dynamics* **233**(3), 1064–1075.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, A.D., Çatalyürek, Ü. V, Toland, A.E., Enright, A., Pringle, T.H., Zahler, A.M., Haussler, D., Erdmann, J., Hengstenberg, C., Ouwehand, W., Samani, N., Schunkert, H., Munzel, T., Lackner, K., Cambien, F., Goodall, A., Tiret, L., Blakenberg, S., and Tregouet, D.-A., 2002. The Human Genome Browser at UCSC. *Genome Research* **12**(6), 996–1006.

Kibar, Z., Capra, V., and Gros, P., 2007. Toward understanding the genetic basis of neural tube defects. *Clinical Genetics* **71**(4), 295–310.

Kietzman, H.W., Everson, J.L., Sulik, K.K., and Lipinski, R.J., 2014. The Teratogenic Effects of Prenatal Ethanol Exposure Are Exacerbated by Sonic Hedgehog or Gli2 Haploinsufficiency in the Mouse. *PLoS ONE* **9**(2), e89448.

Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff-Papadimitriou, E., Kuhl, D., Bito, H., Worley, P.F., Kreiman, G., and Greenberg, M.E., 2010. Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**(7295), 182–7.

Kinder, S.J., Tsang, T.E., Wakamiya, M., Sasaki, H., Behringer, R.R., Nagy, A., and Tam, P.P., 2001. The organizer of the mouse gastrula is composed of a dynamic population of progenitor cells for the axial mesoderm. *Development* **128**(18), 3623–3634.

Kingsley, D.M., Bland, A.E., Grubber, J.M., Marker, P.C., Russell, L.B., Copeland, N.G., and Jenkins, N.A., 1992. The mouse short ear skeletal morphogenesis locus is associated with defects in a bone morphogenetic member of the TGF beta superfamily. *Cell* **71**(3), 399–410.

Kinzler, K.W., and Vogelstein, B., 1990. The GLI gene encodes a nuclear protein which binds specific sequences in the human genome. *Molecular and Cellular Biology* **10**(2), 634–42.

Klingensmith, J., Ang, S.-L., Bachiller, D., and Rossant, J., 1999. Neural Induction and Patterning in the Mouse in the Absence of the Node and Its Derivatives. *Developmental Biology* **216**(2), 535–549.

Koch, F., Fenouil, R., Gut, M., Cauchy, P., Albert, T.K., Zacarias-Cabeza, J., Spicuglia, S., de la Chapelle, a L., Heidemann, M., Hintermair, C., Eick, D., Gut, I., Ferrier, P., and Andrau, J.-C., 2011. Transcription initiation platforms and GTF recruitment at tissue-specific enhancers and promoters. *Nature Structural & Molecular Biology* **18**(8), 956-U124.

Korchynskyi, O., and Ten Dijke, P., 2002. Identification and functional characterization of distinct critically important bone morphogenetic protein-specific response elements in the Id1 promoter. *Journal of Biological Chemistry* **277**(7), 4883–4891.

Koyabu, Y., Nakata, K., Mizugishi, K., Aruga, J., and Mikoshiba, K., 2001. Physical and Functional Interactions between Zic and Gli Proteins. *Journal of Biological Chemistry* **276**(10), 6889–6892.

Krauss, R.S., 2007. Holoprosencephaly: new models, new insights. *Expert Reviews in Molecular Medicine* **9**(26), 1–17.

Krebs, L.T., Iwai, N., Nonaka, S., Welsh, I.C., Lan, Y., Jiang, R., Saijoh, Y., O'Brien, T.P., Hamada, H., and Gridley, T., 2003. Notch signaling regulates left-right asymmetry determination by inducing Nodal expression. *Genes & Development* **17**(10), 1207–12.

Kubo, A., 2004. Development of definitive endoderm from embryonic stem cells in culture. *Development* **131**(7), 1651–1662.

Kuersten, S., and Goodwin, E.B., 2003. The power of the 3' UTR: translational control and development. *Nature reviews. Genetics* **4**(8), 626–637.

Kulkarni, M., Ozgur, S., and Stoecklin, G., 2010. On track with P-bodies. *Biochemical Society transactions* **38**(Pt 1), 242–51.

Kulkarni, M.M., and Arnosti, D.N., 2003. Information display by transcriptional enhancers. *Development* **130**(26).

Kumar, A., Novoselov, V., Celeste, A.J., Wolfman, N.M., Ten Dijke, P., and Kuehn, M.R., 2001. Nodal signaling uses activin and transforming growth factor-beta receptor-regulated Smads. *Journal of Biological Chemistry* **276**(1), 656–661.

Kuras, L., Borggrefe, T., and Kornberg, R.D., 2003. Association of the Mediator complex with enhancers of active genes. *Proceedings of the National Academy of Sciences of the United States of America* **100**(24), 13887–91.

Lacbawan, F., Solomon, B.D., Roessler, E., El-Jaick, K., Domene, S., Velez, J.I., Zhou, N., Hadley, D., Balog, J.Z., Long, R., Fryer, A., Smith, W., Omar, S., McLean, S.D., Clarkson, K., Lichty, A., Clegg, N.J., Delgado, M.R., Levey, E., Stashinko, E., Potocki, L., VanAllen, M.I., Clayton-Smith, J., Donnai, D., Bianchi, D.W., Juliusson, P.B., Njolstad, P.R., Brunner, H.G., Carey, J.C., Hehr, U., Musebeck, J., Wieacker, P.F., Postra, A., Hennekam, R.C.M., van den Boogaard, M.-J.H., van Haeringen, A., Paulussen, A., Herbergs, J., Schrander-Stumpel, C.T.R.M., Janecke, A.R., Chitayat, D., Hahn, J., McDonald-McGinn, D.M., Zackai, E.H., Dobyns, W.B., and Muenke, M., 2009. Clinical spectrum of SIX3-associated mutations in holoprosencephaly: correlation between genotype, phenotype and function. *Journal of Medical Genetics* **46**(6), 389–398.

Lam, M.T.Y., Cho, H., Lesch, H.P., Gosselin, D., Heinz, S., Tanaka-Oishi, Y., Benner, C., Kaikkonen, M.U., Kim, A.S., Kosaka, M., Lee, C.Y., Watt, A., Grossman, T.R., Rosenfeld,

M.G., Evans, R.M., and Glass, C.K., 2013. Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. *Nature* **498**(7455), 511–515.

Lam, M.T.Y., Li, W., Rosenfeld, M.G., and Glass, C.K., 2014. Enhancer RNAs and regulated transcriptional programs. *Trends in Biochemical Sciences* **39**(4), 170–182.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S., 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**(3), R25.

Lawson, K.A., Dunn, N.R., Roelen, B.A., Zeinstra, L.M., Davis, A.M., Wright, C. V, Korving, J.P., and Hogan, B.L., 1999. Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes & Development* **13**(4), 424–36.

Lawson, K.A., Meneses, J.J., and Pedersen, R.A., 1991. Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* **113**(3), 891–911.

Le Dréau, G., and Martí, E., 2012. Dorsal-ventral patterning of the neural tube: A tale of three signals. *Developmental Neurobiology* **72**(12), 1471–1481.

Lee, J.D., and Anderson, K. V., 2008. Morphogenesis of the node and notochord: the cellular basis for the establishment and maintenance of left-right asymmetry in the mouse. *Developmental Dynamics* **237**(12), 3464–76.

Lettice, L.A., Heaney, S.J.H., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D.K., Elgar, G., Hill, R.E., and de Graaff, E., 2003. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics* **12**(14), 1725–35.

Levine, M., and Tjian, R., 2003. Transcription regulation and animal diversity. *Nature* **424**(6945), 147–151.

Lewis, A.J., Simon, E.M., Barkovich, A.J., Clegg, N.J., Delgado, M.R., Levey, E., and Hahn, J.S., 2002. Middle interhemispheric variant of holoprosencephaly: a distinct cliniconeuroradiologic subtype. *Neurology* **59**(12), 1860–5.

Lewis, B.P., Burge, C.B., and Bartel, D.P., 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**(1), 15–20.

Lewis, S.L., and Tam, P.P.L., 2006. Definitive endoderm of the mouse embryo: Formation, cell fates, and morphogenetic function. *Developmental Dynamics* **235**(9), 2315–2329.

Li, M., Defren, J., and Brewer, G., 2013. Hsp27 and F-box protein β-TrCP promote degradation of mRNA decay factor AUF1. *Molecular and Cellular Biology* **33**(11), 2315–26.

Li, W., Notani, D., Ma, Q., Tanasa, B., Nunez, E., Chen, A.Y., Merkurjev, D., Zhang, J., Ohgi, K.A., Song, X., Oh, S., Kim, H.-S., Glass, C.K., and Rosenfeld, M.G., 2013. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**(7455), 516–20.

Li, X.-J., Zhang, X., Johnson, M.A., Wang, Z.-B., LaVaute, T., and Zhang, S.-C., 2009. Coordination of sonic hedgehog and Wnt signaling determines ventral and dorsal telencephalic neuron types from human embryonic stem cells. *Development* **136**(23), 4055–4063.

Liem, K.F., Ashe, A., He, M., Satir, P., Moran, J., Beier, D.R., Wicking, C., and Anderson, K. V., 2012. The IFT-A complex regulates Shh signaling through cilia structure and membrane protein trafficking. *The Journal of Cell Biology* **197**(6), 789–800.

Liem, K.F., Tremml, G., Roelink, H., and Jessell, T.M., 1995. Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm. *Cell* **82**(6), 969–79.

Lin, Y.-C., Boone, M., Meuris, L., Lemmens, I., Van Roy, N., Soete, A., Reumers, J., Moisse, M., Plaisance, S., Drmanac, R., Chen, J., Speleman, F., Lambrechts, D., Van de Peer, Y., Tavernier, J., and Callewaert, N., 2014. Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. *Nature communications* **5**(11), 4767.

Lin, Y.C., Benner, C., Mansson, R., Heinz, S., Miyazaki, K., Miyazaki, M., Chandra, V., Bossen, C., Glass, C.K., and Murre, C., 2012. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nature immunology* **13**(12), 1196–204.

Liu, H., Han, H., Li, J., and Wong, L., 2003. An in-silico method for prediction of polyadenylation signals in human sequences. *Genome Informatics: International Conference on Genome Informatics* **14**, 84–93.

Liu, Y., Festing, M., Thompson, J.C., Hester, M., Rankin, S., El-Hodiri, H.M., Zorn, A.M., and Weinstein, M., 2004. Smad2 and Smad3 coordinately regulate craniofacial and endodermal development. *Developmental Biology* **270**(2), 411–426.

Livak, K.J., and Schmittgen, T.D., 2001. Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods* **25**(4), 402–408.

Loflin, P., Chen, C.-Y.A., and Shyu, A. Bin, 1999. Unraveling a cytoplasmic role for hnRNP d in the in vivo mRNA destabilization directed by the AU-rich element. *Genes & Development* **13**(14), 1884–1897.

Loots, G.G., and Ovcharenko, I., 2007. ECRbase: Database of evolutionary conserved regions, promoters, and transcription factor binding sites in vertebrate genomes. *Bioinformatics* **23**(1), 122–124.

Lowe, L.A., Supp, D.M., Sampath, K., Yokoyama, T., Wright, C.V.E., Potter, S.S., Overbeek, P., and Kuehn, M.R., 1996. Conserved left-right asymmetry of nodal expression and alterations in murine situs inversus. *Nature* **381**(6578), 158–161.

Lu, W., 2001. Comparison of Pkd1-targeted mutants reveals that loss of polycystin-1 causes cystogenesis and bone defects. *Human Molecular Genetics* **10**(21), 2385–2396.

Luo, G., Hofmann, C., Bronckers, A.L., Sohocki, M., Bradley, A., and Karsenty, G., 1995. BMP-7 is an inducer of nephrogenesis, and is also required for eye development and skeletal patterning. *Genes & Development* **9**(22), 2808–20.

Madabhushi, M., and Lacy, E., 2011. Anterior visceral endoderm directs ventral morphogenesis and placement of head and heart via BMP2 expression. *Developmental cell* **21**(5), 907–19.

Mahon, M.J., 2011. Vectors bicistronically linking a gene of interest to the SV40 large T antigen in combination with the SV40 origin of replication enhance transient protein expression and luciferase reporter activity. *BioTechniques* **51**(2), 119–28.

Maisonneuve, C., Guilleret, I., Vick, P., Weber, T., Andre, P., Beyer, T., Blum, M., and Constam, D.B., 2009. Bicaudal C, a novel regulator of Dvl signaling abutting RNA-processing bodies, controls cilia orientation and leftward flow. *Development* **136**(17), 3019–30.

Mannervik, M., Nibu, Y., Zhang, H., and Levine, M., 1999. Transcriptional coregulators in development. *Science* **284**(5414), 606–609.

Marcorelles, P., and Laquerriere, A., 2010. Neuropathology of holoprosencephaly. *American Journal of Medical Genetics Part C: Seminars in Medical Genetics* **154C**(1), 109–119.

Marques, S., Borges, A.C., Silva, A.C., Freitas, S., Cordenonsi, M., and Belo, J.A., 2004. The activity of the Nodal antagonist Cerl-2 in the mouse node is required for correct L/R body axis. *Genes & Development* **18**(19), 2342–2347.

Martín, C., Bueno, D., Alonso, M.I., Moro, J.A., Callejo, S., Parada, C., Martín, P., Carnicero, E., and Gato, A., 2006. FGF2 plays a key role in embryonic cerebrospinal fluid trophic properties over chick embryo neuroepithelial stem cells. *Developmental biology* **297**(2), 402–16.

Martin, D.M., Skidmore, J.M., Fox, S.E., Gage, P.J., and Camper, S.A., 2002. Pitx2 distinguishes subtypes of terminally differentiated neurons in the developing mouse neuroepithelium. *Developmental Biology* **252**(1), 84–99.

Martin, G.R., 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **78**(12), 7634–8.

Maston, G.A., Evans, S.K., and Green, M.R., 2006. Transcriptional regulatory elements in the human genome. *Annual review of genomics and human genetics* **7**, 29–59.

Mátés, L., 2011. Rodent transgenesis mediated by a novel hyperactive Sleeping Beauty transposon system. *Methods in molecular biology (Clifton, N.J.)* **738**, 87–99.

Mátés, L., Chuah, M.K.L., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D.P., Schmitt, A., Becker, K., Matrai, J., Ma, L., Samara-Kuko, E., Gysemans, C., Pryputniewicz, D., Miskey, C., Fletcher, B., VandenDriessche, T., Ivics, Z., and Izsvák, Z., 2009. Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nature Genetics* **41**(6), 753–761.

Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C., Chou, A., Ienasescu, H., Lim, J., Shyr, C., Tan, G., Zhou, M., Lenhard, B., Sandelin, A., and Wasserman, W.W., 2014. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic Acids Research* **42**(Database issue), D142--7.

Matsunaga, E., and Shiota, K., 1977. Holoprosencephaly in human embryos: Epidemiologic studies of 150 cases. *Teratology* **16**(3), 261–272.

Matthaei, K.I., 2007. Genetically manipulated mice: a powerful tool with unsuspected caveats. *The Journal of Physiology* **582**(2), 481–488.

Matys, V., Kel-Margoulis, O. V, Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K., Voss, N., Stegmaier, P., Lewicki-Potapov, B., Saxel, H., Kel, A.E., and Wingender, E., 2006. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Research* **34**(Database issue), D108--10.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutyavin, T., Stehling-Sun, S., Johnson, A.K., Canfield, T.K., Giste, E., Diegel, M., Bates, D., Hansen, R.S., Neph, S., Sabo, P.J., Heimfeld, S., Raubitschek, A., Ziegler, S., Cotsapas, C., Sotoodehnia, N., Glass, I.,

Sunyaev, S.R., Kaul, R., and Stamatoyannopoulos, J.A., 2012. Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science* **337**(6099), 1190–1195.

McGrath, J., Somlo, S., Makova, S., Tian, X., and Brueckner, M., 2003. Two populations of node monocilia initiate left-right asymmetry in the mouse. *Cell* **114**(1), 61–73.

McGray, A.J.R., Gingerich, T., Petrik, J.J., and Lamarre, J., 2011. Regulation of thrombospondin-1 expression through AU-rich elements in the 3'UTR of the mRNA. *Cellular & molecular biology letters* **16**(1), 55–68.

McLean, C., and Bejerano, G., 2008. Dispensability of mammalian DNA. *Genome Research* **18**(11), 1743–1751.

Megason, S.G., and McMahon, A.P., 2002. A mitogen gradient of dorsal midline Wnts organizes growth in the CNS. *Development* **129**(9), 2087–98.

Meno, C., Saijoh, Y., Fujii, H., Ikeda, M., Yokoyama, T., Yokoyama, M., Toyoda, Y., and Hamada, H., 1996. Left-right asymmetric expression of the TGF beta-family member lefty in mouse embryos. *Nature* **381**(6578), 151–155.

Meno, C., Shimono, A., Saijoh, Y., Yashiro, K., Mochida, K., Ohishi, S., Noji, S., Kondoh, H., and Hamada, H., 1998. Lefty-1 is required for left-right determination as a regulator of Lefty-2 and Nodal. *Cell* **94**(3), 287–297.

Mercier, S., Dubourg, C., Garcelon, N., Campillo-Gimenez, B., Gicquel, I., Belleguic, M., Ratié, L., Pasquier, L., Loget, P., Bendavid, C., Jaillard, S., Rochard, L., Quelin, C., Dupé, V., David, V., and Odent, S., 2011. New findings for phenotype-genotype correlations in a large European series of holoprosencephaly cases. *Journal of Medical Genetics* **48**(11), 752–760.

Merika, M., and Thanos, D., 2001. Enhanceosomes. *Current opinion in genetics & development* **11**(2), 205–208.

Meyer, M.B., Benkusky, N.A., and Pike, J.W., 2015. Selective Distal Enhancer Control of the Mmp13 Gene Identified through Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) Genomic Deletions. *The Journal of biological chemistry* **290**(17), 11093–107.

Mianné, J., Codner, G., Caulder, A., Fell, R., Hutchison, M., King, R., Stewart, M.E., Wells, S., and Teboul, L., 2017. Analysing the outcome of CRISPR-aided genome editing in embryos: Screening, genotyping and quality control. *Methods* **121**–**122**.

Michalova, E., Vojtesek, B., and Hrstka, R., 2013. Impaired pre-mRNA processing and altered architecture of 3' untranslated regions contribute to the development of human disorders. *International Journal of Molecular Sciences* **14**(8), 15681–15694.

Ming, J.E., Kaupas, M.E., Roessler, E., Brunner, H.G., Golabi, M., Tekin, M., Stratton, R.F., Sujansky, E., Bale, S.J., and Muenke, M., 2002. Mutations in PATCHED-1, the receptor for SONIC HEDGEHOG, are associated with holoprosencephaly. *Human Genetics* **110**(4), 297–301.

Mizugishi, K., Aruga, J., Nakata, K., and Mikoshiba, K., 2001. Molecular properties of Zic proteins as transcriptional regulators and their relationship to GLI proteins. *Journal of Biological Chemistry* **276**(3), 2180–2188.

Mizugishi, K., Hatayama, M., Tohmonda, T., Ogawa, M., Inoue, T., Mikoshiba, K., and Aruga, J.,

2004. Myogenic repressor I-mfa interferes with the function of Zic family proteins. *Biochemical and Biophysical Research Communications* **320**(1), 233–240.

Monaghan, A.P., Kaestner, K.H., Grau, E., and Schütz, G., 1993. Postimplantation expression patterns indicate a role for the mouse forkhead/HNF-3 alpha, beta and gamma genes in determination of the definitive endoderm, chordamesoderm and neuroectoderm. *Development* **119**(3), 567–78.

Moorman, A.F.M., and Christoffels, V.M., 2003. Cardiac chamber formation: development, genes, and evolution. *Physiological Reviews* **83**(4), 1223–1267.

Morel, V., Lecourtois, M., Massiani, O., Maier, D., Preiss, A., Schweisguth, F., and al., et, 2001. Transcriptional repression by suppressor of hairless involves the binding of a hairless-dCtBP complex in Drosophila. *Current biology : CB* **11**(10), 789–92.

Morikawa, M., Koinuma, D., Tsutsumi, S., Vasilaki, E., Kanki, Y., Heldin, C.H., Aburatani, H., and Miyazono, K., 2011. ChIP-seq reveals cell type-specific binding patterns of BMP-specific Smads and a novel binding motif. *Nucleic Acids Research* **39**(20), 8712–8727.

Mouden, C., Dubourg, C., Carré, W., Rose, S., Quelin, C., Akloul, L., Viot, G., Salhi, H., Darnault, P., Odent, S., Dupé, V., and David, V., 2016. Complex mode of inheritance in holoprosencephaly revealed by whole exome sequencing. *Clinical Genetics* **89**(6), 659–68.

Murdoch, J.N., and Copp, A.J., 2010. The relationship between sonic Hedgehog signaling, cilia, and neural tube defects. *Birth defects research. Part A, Clinical and molecular teratology* **88**(8), 633–52.

Nagai, T., Aruga, J., Minowa, O., Sugimoto, T., Ohno, Y., Noda, T., and Mikoshiba, K., 2000. Zic2 regulates the kinetics of neurulation. *Proceedings of the National Academy of Sciences of the United States of America* **97**(4), 1618–1623.

Nagai, T., Aruga, J., Takada, S., Günther, T., Spörle, R., Schughart, K., and Mikoshiba, K., 1997. The expression of the mouse Zic1, Zic2, and Zic3 gene suggests an essential role for Zic genes in body pattern formation. *Developmental Biology* **182**(2), 299–313.

Nakamura, T., Saito, D., Kawasumi, A., Shinohara, K., Asai, Y., Takaoka, K., Dong, F., Takamatsu, A., Belo, J.A., Mochizuki, A., and Hamada, H., 2012. Fluid flow and interlinked feedback loops establish left–right asymmetric decay of Cerl2 mRNA. *Nature Communications* **3**, 1322.

Nakanishi, T., Kuroiwa, A., Yamada, S., Isotani, A., Yamashita, A., Tairaka, A., Hayashi, T., Takagi, T., Ikawa, M., Matsuda, Y., and Okabe, M., 2002. FISH Analysis of 142 EGFP Transgene Integration Sites into the Mouse Genome. *Genomics* **80**(6), 564–574.

Nakayama, J., Kinugasa, H., Ohto, T., Tanaka, R., Nakayama, T., Noguchi, E., Arinami, T., and Iwasaki, N., 2016. Monozygotic twins with de novo ZIC2 gene mutations discordant for the type of holoprosencephaly. *Neurology* **86**(15), 1456–1458.

Nanni, L., Ming, J.E., Bocian, M., Steinhaus, K., Bianchi, D.W., Die-Smulders, C., Giannotti, A., Imaizumi, K., Jones, K.L., Campo, M.D., Martin, R. a, Meinecke, P., Pierpont, M.E., Robin, N.H., Young, I.D., Roessler, E., and Muenke, M., 1999. The mutational spectrum of the sonic hedgehog gene in holoprosencephaly: SHH mutations cause a significant proportion of autosomal dominant holoprosencephaly. *Human Molecular Genetics* **8**(13), 2479–2488.

Nichols, J., and Smith, A., 2009. Naive and Primed Pluripotent States. *Cell Stem Cell* **4**(6), 487–492.

Nikonova, A.S., Plotnikova, O. V, Serzhanova, V., Efimov, A., Bogush, I., Cai, K.Q., Hensley, H.H., Egleston, B.L., Klein-Szanto, A., Seeger-Nukpezah, T., and Golemis, E.A., 2014. Nedd9 restrains renal cystogenesis in Pkd1-/- mice. *Proceedings of the National Academy of Sciences of the United States of America* **111**(35), 12859–64.

Nonaka, S., Shiratori, H., Saijoh, Y., and Hamada, H., 2002. Determination of left-right patterning of the mouse embryo by artificial nodal flow. *Nature* **418**(6893), 96–99.

Nonaka, S., Tanaka, Y., Okada, Y., Takeda, S., Harada, A., Kanai, Y., Kido, M., and Hirokawa, N., 1998. Randomization of left-right asymmetry due to loss of nodal cilia generating leftward flow of extraembryonic fluid in mice lacking KIF3B motor protein. *Cell* **95**(6), 829–837.

Nonaka, S., Yoshiba, S., Watanabe, D., Ikeuchi, S., Goto, T., Marshall, W.F., and Hamada, H., 2005. De novo formation of left-right asymmetry by posterior tilt of nodal cilia. *PLoS Biology* **3**(8).

Norris, D.P., 2012. Cilia, calcium and the basis of left-right asymmetry. *BMC Biology* **10**, 102.

Norris, D.P., Brennan, J., Bikoff, E.K., and Robertson, E.J., 2002. The Foxh1-dependent autoregulatory enhancer controls the level of Nodal signals in the mouse embryo. *Development* **129**(14), 3455–3468.

Norris, D.P., and Grimes, D.T., 2012. Cilia discern left from right. *Science* **338**(6104), 206–7.

Ocbina, P.J.R., Tuson, M., and Anderson, K. V., 2009. Primary cilia are not required for normal canonical Wnt signaling in the mouse embryo. *PLoS ONE* **4**(8).

Odent, S., Le Marec, B., Munnich, A., Le Merrer, M., and Bonaïti-Pellié, C., 1998. Segregation analysis in nonsyndromic holoprosencephaly. *American Journal of Medical Genetics* **77**(2), 139–43.

Odenwald, W.F., Rasband, W., Kuzin, A., and Brody, T., 2005. EVOPRINTER, a multigenomic comparative tool for rapid identification of functionally important DNA. *Proceedings of the National Academy of Sciences of the United States of America* **102**(41), 14700–14705.

Ohkubo, Y., Chiang, C., and Rubenstein, J.L.R., 2002. Coordinate regulation and synergistic actions of BMP4, SHH and FGF8 in the rostral prosencephalon regulate morphogenesis of the telencephalic and optic vesicles. *Neuroscience* **111**(1), 1–17.

Okada, T., Okumura, Y., Motoyama, J., and Ogawa, M., 2008. FGF8 signaling patterns the telencephalic midline by regulating putative key factors of midline development. *Developmental Biology* **320**(1), 92–101.

Okada, Y., Nonaka, S., Tanaka, Y., Saijoh, Y., Hamada, H., and Hirokawa, N., 1999. Abnormal nodal flow precedes situs inversus in iv and inv mice. *Molecular Cell* **4**(4), 459–468.

Okada, Y., Takeda, S., Tanaka, Y., Belmonte, J.C.I., and Hirokawa, N., 2005. Mechanism of nodal flow: A conserved symmetry breaking event in left-right axis determination. *Cell* **121**(4), 633–644.

Ong, C.-T., and Corces, V.G., 2011. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature reviews. Genetics* **12**(4), 283–293.

Orioli, I.M., and Castilla, E.E., 2010. Epidemiology of holoprosencephaly: Prevalence and risk factors. *American Journal of Medical Genetics, Part C: Seminars in Medical Genetics* **154**(1), 13–21.

Orkin, S.H., Cheng, T.C., Antonarakis, S.E., and Kazazian, H.H., 1985. Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene. *The EMBO journal* **4**(2), 453–456.

Palstra, R.J., and Grosveld, F., 2012. Transcription factor binding at enhancers: Shaping a genomic regulatory landscape in flux. *Frontiers in Genetics* **3**(SEP).

Palstra, R.J., Tolhuis, B., Splinter, E., Nijmeijer, R., Grosveld, F., and de Laat, W., 2003. The beta-globin nuclear compartment in development and erythroid differentiation. *Nature Genetics* **35**(2), 190–194.

Pan, C., Kumar, C., Bohl, S., Klingmueller, U., and Mann, M., 2009. Comparative Proteomic Phenotyping of Cell Lines and Primary Cells to Assess Preservation of Cell Type-specific Functions. *Molecular & Cellular Proteomics* **8**(3), 443–450.

Panchision, D.M., Pickel, J.M., Studer, L., Lee, S.H., Turner, P.A., Hazel, T.G., and McKay, R.D., 2001. Sequential actions of BMP receptors control neural precursor cell production and fate. *Genes & Development* **15**(16), 2094–2110.

Panne, D., 2008. The enhanceosome. *Current Opinion in Structural Biology* **18**(2), 236–42.

Paraskevopoulou, M.D., Georgakilas, G., Kostoulas, N., Vlachos, I.S., Vergoulis, T., Reczko, M., Filippidis, C., Dalamagas, T., and Hatzigeorgiou, A.G., 2013. DIANA-microT web server v5.0: service integration into miRNA functional analysis workflows. *Nucleic Acids Research* **41**(Web Server issue).

Parker, R., and Sheth, U., 2007. P Bodies and the Control of mRNA Translation and Degradation. *Molecular Cell* **25**(5), 635–646.

Parkhomchuk, D., Borodina, T., Amstislavskiy, V., Banaru, M., Hallen, L., Krobitsch, S., Lehrach, H., and Soldatov, A., 2009. Transcriptome analysis by strand-specific sequencing of complementary DNA. *Nucleic Acids Research* **37**(18).

Parr, B.A., Shea, M.J., Vassileva, G., and McMahon, A.P., 1993. Mouse Wnt genes exhibit discrete domains of expression in the early embryonic CNS and limb buds. *Development* **119**(1), 247–61.

Paulussen, A.D.C., Schrander-Stumpel, C.T., Tserpelis, D.C.J., Spee, M.K.M., Stegmann, A.P.A., Mancini, G.M., Brooks, A.S., Collée, M., Maat-Kievit, A., Simon, M.E.H., van Bever, Y., Stolte-Dijkstra, I., Kerstjens-Frederikse, W.S., Herkert, J.C., van Essen, A.J., Lichtenbelt, K.D., van Haeringen, A., Kwee, M.L., Lachmeijer, A.M.A., Tan-Sindhunata, G.M.B., van Maarle, M.C., Arens, Y.H.J.M., Smeets, E.E.J.G.L., de Die-Smulders, C.E., Engelen, J.J.M., Smeets, H.J., and Herbergs, J., 2010. The unfolding clinical spectrum of holoprosencephaly due to mutations in SHH, ZIC2, SIX3 and TGIF genes. *European Journal of Human Genetics* **18**(9), 999–1005.

Pavletich, N.P., and Pabo, C.O., 1993. Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science* **261**(5129), 1701–7.

Payankaulam, S., Li, L.M., and Arnosti, D.N., 2010. Transcriptional repression: Conserved and evolved features. *Current Biology* **20**(17).

Paz, I., Kosti, I., Ares, M., Cline, M., and Mandel-Gutfreund, Y., 2014. RBPmap: A web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Research* **42**(W1).

Pearce, J.J., Penny, G., and Rossant, J., 1999. A mouse cerberus/Dan-related gene family. *Developmental biology* **209**(1), 98–110.

Peng, S.S., Chen, C.Y., Xu, N., and Shyu, A. Bin, 1998. RNA stabilization by the AU-rich element binding protein, HuR, an ELAV protein. *The EMBO journal* **17**(12), 3461–70.

Pennekamp, P., Karcher, C., Fischer, A., Schweickert, A., Skryabin, B., Horst, J., Blum, M., and Dworniczak, B., 2002. The ion channel polycystin-2 is required for left-right axis determination in mice. *Current Biology* **12**(11), 938–943.

Perea-Gomez, A., Vella, F.D.J., Shawlot, W., Oulad-Abdelghani, M., Chazaud, C., Meno, C., Pfister, V., Chen, L., Robertson, E., Hamada, H., Behringer, R.R., and Ang, S.-L., 2002. Nodal antagonists in the anterior visceral endoderm prevent the formation of multiple primitive streaks. *Developmental cell* **3**(5), 745–56.

Petryk, A., Graf, D., and Marcucio, R., 2015. Holoprosencephaly: Signaling interactions between the brain and the face, the environment and the genes, and the phenotypic variability in animal models and humans. *Wiley Interdisciplinary Reviews: Developmental Biology* **4**(1), 17–32.

Pfaffl, M.W., Tichopad, A., Prgomet, C., and Neuvians, T.P., 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. *Biotechnology letters* **26**(6), 509–15.

Pfister, S., Steiner, K.A., and Tam, P.P.L., 2007. Gene expression pattern and progression of embryogenesis in the immediate post-implantation period of mouse development. *Gene expression patterns : GEP* **7**(5), 558–73.

Pillai, R.S., Bhattacharyya, S.N., Artus, C.G., Zoller, T., Cougot, N., Basyuk, E., Bertrand, E., and Filipowicz, W., 2005. Inhibition of translational initiation by Let-7 MicroRNA in human cells. *Science* **309**(5740), 1573–1576.

Pineda-Alvarez, D.E., Roessler, E., Hu, P., Srivastava, K., Solomon, B.D., Siple, C.E., Fan, C.-M., and Muenke, M., 2012. Missense substitutions in the GAS1 protein present in holoprosencephaly patients reduce the affinity for its ligand, SHH. *Human Genetics* **131**(2), 301–10.

Placzek, M., and Briscoe, J., 2005. The floor plate: multiple cells, multiple signals. *Nature Reviews Neuroscience* **6**(3), 230–240.

Poitras, L., Yu, M., Lesage-Pelletier, C., Macdonald, R.B., Gagné, J.-P., Hatch, G., Kelly, I., Hamilton, S.P., Rubenstein, J.L.R., Poirier, G.G., and Ekker, M., 2010. An SNP in an ultraconserved regulatory element affects Dlx5/Dlx6 regulation in the forebrain. *Development* **137**(18), 3089–97.

Pourebrahim, R., Houtmeyers, R., Ghogomu, S., Janssens, S., Thelie, A., Tran, H.T., Langenberg, T., Vleminckx, K., Bellefroid, E., Cassiman, J.J., and Tejpar, S., 2011. Transcription factor Zic2 inhibits Wnt/Beta-catenin protein signaling. *Journal of Biological Chemistry* **286**(43), 37732–37740.

Prestridge, D.S., 1995. Predicting Pol II Promoter Sequences using Transcription Factor Binding Sites. *Journal of Molecular Biology* **249**(5), 923–932.

Quintana, L., and Sharpe, J., 2011. Preparation of mouse embryos for optical projection tomography imaging. *Cold Spring Harbor Protocols* **2011**(6), 664–9.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J., 2011. A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**(7333), 279–83.

Raghavan, A., Ogilvie, R.L., Reilly, C., Abelson, M.L., Raghavan, S., Vasdewani, J., Krathwohl, M., and Bohjanen, P.R., 2002. Genome-wide analysis of mRNA decay in resting and activated primary human T lymphocytes. *Nucleic Acids Research* **30**(24), 5529–5538.

Rahimov, F., Marazita, M.L., Visel, A., Cooper, M.E., Hitchler, M.J., Rubini, M., Domann, F.E., Govil, M., Christensen, K., Bille, C., Melbye, M., Jugessur, A., Lie, R.T., Wilcox, A.J., Fitzpatrick, D.R., Green, E.D., Mossey, P.A., Little, J., Steegers-Theunissen, R.P., Pennacchio, L.A., Schutte, B.C., and Murray, J.C., 2008. Disruption of an AP-2alpha binding site in an IRF6 enhancer is associated with cleft lip. *Nature Genetics* **40**(11), 1341–7.

Reczko, M., Maragkakis, M., Alexiou, P., Grosse, I., and Hatzigeorgiou, A.G., 2012. Functional microRNA targets in protein coding sequences. *Bioinformatics* **28**(6), 771–6.

Ribatti, D., Nico, B., Vacca, a, Roncali, L., and Presta, M., 1999. Endogenous and exogenous fibroblast growth factor-2 modulate wound healing in the chick embryo chorioallantoic membrane. *Angiogenesis* **3**(1), 89–95.

Ribeiro, L.A., Quiezi, R.G., Nascimento, A., Bertolacini, C.P., and Richieri-Costa, A., 2010. Holoprosencephaly and holoprosencephaly-like phenotype and GAS1 DNA sequence changes: Report of four Brazilian patients. *American Journal of Medical Genetics Part A* **152A**(7), 1688–1694.

Ribeiro, L.A., Roessler, E., Hu, P., Pineda-Alvarez, D.E., Zhou, N., Jones, M., Chandrasekharappa, S., Richieri-Costa, A., and Muenke, M., 2012. Comparison of mutation findings in ZIC2 between microform and classical holoprosencephaly in a Brazilian cohort. *Birth Defects Research Part A - Clinical and Molecular Teratology* **94**(11), 912–917.

Richardson, L., Venkataraman, S., Stevenson, P., Yang, Y., Moss, J., Graham, L., Burton, N., Hill, B., Rao, J., Baldock, R. a., and Armit, C., 2014. EMAGE mouse embryo spatial gene expression database: 2014 update. *Nucleic Acids Research* **42**(D1), 1–10.

Rivera-Pérez, J.A., and Magnuson, T., 2005. Primitive streak formation in mice is preceded by localized activation of Brachyury and Wnt3. *Developmental Biology* **288**(2), 363–371.

Rivera-Perez, J., and Hadjantonakis, A.K., 2014. The Dynamics of Morphogenesis in the Early Mouse Embryo. *Cold Spring Harbor Perspectives in Biology* **257**(5), 1–28.

Robb, L., and Tam, P.P.L., 2004. Gastrula organiser and embryonic patterning in the mouse. *Seminars in Cell & Developmental Biology* **15**(5), 543–554.

Robertson, E.J., Norris, D.P., Brennan, J., and Bikoff, E.K., 2003. Control of early anterior-posterior patterning in the mouse embryo by TGF-beta signalling. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **358**(1436), 1351–7; discussion 1357.

Roessler, E., Belloni, E., Gaudenz, K., Jay, P., Berta, P., Scherer, S.W., Tsui, L.-C., and Muenke, M., 1996. Mutations in the human Sonic Hedgehog gene cause holoprosencephaly. *Nature Genetics* **14**(3), 357–360.

Roessler, E., Hu, P., Hong, S.-K., Srivastava, K., Carrington, B., Sood, R., Petrykowska, H.M., Elnitski, L., Ribeiro, L.A., Richieri-Costa, A., Feldman, B., Odenwald, W.F., and Muenke, M., 2012a. Unique alterations of an ultraconserved non-coding element in the 3'UTR of ZIC2 in holoprosencephaly. *PLoS ONE* **7**(7), 3–8.

Roessler, E., Lacbawan, F., Dubourg, C., Paulussen, A.D.C., Odent, S., David, V., and Muenke, M., 2009a. The Full Spectrum of Holoprosencephaly-Associated Mutations within the

ZIC2 Gene in Humans Predicts Loss-of-Function as the Predominant Disease Mechanism. *Human Mutation* **30**(4).

Roessler, E., Ma, Y., Ouspenskaia, M. V, Lacbawan, F., Bendavid, C., Dubourg, C., Beachy, P.A., and Muenke, M., 2009b. Truncating loss-of-function mutations of DISP1 contribute to holoprosencephaly-like microform features in humans. *Human Genetics* **125**(4), 393–400.

Roessler, E., and Muenke, M., 2001. Midline and laterality defects: Left and right meet in the middle. *BioEssays* **23**(10), 888–900.

Roessler, E., and Muenke, M., 2010. The molecular genetics of holoprosencephaly. *American Journal of Medical Genetics, Part C: Seminars in Medical Genetics* **154**(1), 52–61.

Roessler, E., Pei, W., Ouspenskaia, M. V, Karkera, J.D., Veléz, J.I., Banerjee-Basu, S., Gibney, G., Lupo, P.J., Mitchell, L.E., Towbin, J.A., Bowers, P., Belmont, J.W., Goldmuntz, E., Baxevanis, A.D., Feldman, B., and Muenke, M., 2009c. Cumulative ligand activity of NODAL mutations and modifiers are linked to human heart defects and holoprosencephaly. *Molecular Genetics and Metabolism* **98**(1–2), 225–234.

Roessler, E., Vélez, J.I., Zhou, N., and Muenke, M., 2012b. Utilizing prospective sequence analysis of SHH, ZIC2, SIX3 and TGIF in holoprosencephaly probands to describe the parameters limiting the observed frequency of mutant gene x gene interactions. *Molecular Genetics and Metabolism* **105**(4), 658–664.

Rosen, B., and Beddington, R.S., 1993. Whole-mount in situ hybridization in the mouse embryo: gene expression in three dimensions. *Trends in Genetics* **9**(5), 162–7.

Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P. a., Guruvadoo, L., Haeussler, M., Harte, R.A., Heitner, S.G., Hickey, G., Hinrichs, A.S., Hubley, R., Karolchik, D., Learned, K., Lee, B.T., Li, C.H., Miga, K.H., Nguyen, N., Paten, B., Raney, B.J., Smit, A.F. a., Speir, M.L., Zweig, A.S., Haussler, D., Kuhn, R.M., and Kent, W.J., 2015. The UCSC Genome Browser database: 2015 update. *Nucleic Acids Research* **43**(Database issue), D670-81.

Rosenbloom, K.R., Sloan, C.A., Malladi, V.S., Dreszer, T.R., Learned, K., Kirkup, V.M., Wong, M.C., Maddren, M., Fang, R., Heitner, S.G., Lee, B.T., Barber, G.P., Harte, R.A., Diekhans, M., Long, J.C., Wilder, S.P., Zweig, A.S., Karolchik, D., Kuhn, R.M., Haussler, D., and Kent, W.J., 2013. ENCODE data in the UCSC Genome Browser: year 5 update. *Nucleic Acids Research* **41**(Database issue), D56-63.

Ruiz i Altaba, A., Prezioso, V.R., Darnell, J.E., and Jessell, T.M., 1993. Sequential expression of HNF-3 beta and HNF-3 alpha by embryonic organizing centers: the dorsal lip/node, notochord and floor plate. *Mechanisms of Development* **44**(2–3), 91–108.

Ryan, A.K., Blumberg, B., Rodriguez-Esteban, C., Yonei-Tamura, S., Tamura, K., Tsukui, T., de la Peña, J., Sabbagh, W., Greenwald, J., Choe, S., Norris, D.P., Robertson, E.J., Evans, R.M., Rosenfeld, M.G., and Izpisúa Belmonte, J.C., 1998. Pitx2 determines left-right asymmetry of internal organs in vertebrates. *Nature* **394**(6693), 545–551.

Sahlén, P., Abdullayev, I., Ramsköld, D., Matskova, L., Rilakovic, N., Lötstedt, B., Albert, T.J., Lundeberg, J., and Sandberg, R., 2015. Genome-wide mapping of promoter-anchored interactions with close to single-enhancer resolution. *Genome Biology* **16**(1), 156.

Sakai-Kato, K., Ishiguro, A., Mikoshiba, K., Aruga, J., and Utsunomiya-Tate, N., 2008. CD spectra show the relational style between Zic-, Gli-, Glis-zinc finger protein and DNA. *Biochimica et Biophysica Acta* **1784**(7–8), 1011–1019.

Sanjana, N.E., Wright, J., Zheng, K., Shalem, O., Fontanillas, P., Joung, J., Cheng, C., Regev, A., and Zhang, F., 2016. High-resolution interrogation of functional elements in the noncoding genome. *doi.org* **34**(9), 167–174.

Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J., 2012. The long-range interaction landscape of gene promoters. *Nature* **489**(7414), 109–13.

Sasaki, H., and Hogan, B.L., 1993. Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo. *Development* **118**(1), 47–59.

Savory, J.G.A., Mansfield, M., Rijli, F.M., and Lohnes, D., 2011. Cdx mediates neural tube closure through transcriptional regulation of the planar cell polarity gene Ptk7. *Development* **138**, 1361–1370.

Schachter, K. a, and Krauss, R.S., 2008. Murine models of holoprosencephaly., 1st ed, Current Topics in Developmental Biology. Elesvier Inc.

Selbach, M., Schwanhäusser, B., Thierfelder, N., Fang, Z., Khanin, R., and Rajewsky, N., 2008. Widespread changes in protein synthesis induced by microRNAs. *Nature* **455**(7209), 58–63.

Serbedzija, G.N., Fraser, S.E., and Bronner-Fraser, M., 1990. Pathways of trunk neural crest cell migration in the mouse embryo as revealed by vital dye labelling. *Development* **108**(4), 605–12.

Sharma, N., Hollensen, A.K., Bak, R.O., Staunstrup, N.H., Schrøder, L.D., and Mikkelsen, J.G., 2012. The Impact of cHS4 Insulators on DNA Transposon Vector Mobilization and Silencing in Retinal Pigment Epithelium Cells. *PLoS ONE* **7**(10), e48421.

Sharova, L. V., Sharov, A.A., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M.S.H., 2009. Database of mRNA Half-Life of 19977 Genes Obtained by DNA Microarray Analysis of Pluripotent and Differentiating Mouse Embryonic Stem Cells Supplementary data. *DNA research* **16**(1), S1.

Shaw, G., and Kamen, R., 1986. A conserved AU sequence from the 3' untranslated region of GM-CSF mRNA mediates selective mRNA degradation. *Cell* **46**(5), 659–667.

Shaw, G., Morse, S., Ararat, M., and Graham, F.L., 2002. Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *The FASEB Journal* **16**(8), 869–71.

Shawlot, W., and Behringer, R.R., 1995. Requirement for Lim1 in head-organizer function. *Nature* **374**(6521), 425–430.

Shimamura, K., and Rubenstein, J.L., 1997. Inductive interactions direct early regionalization of the mouse forebrain. *Development* **124**(14), 2709–18.

Shinohara, K., Kawasumi, A., Takamatsu, A., Yoshiba, S., Botilde, Y., Motoyama, N., Reith, W., Durand, B., Shiratori, H., and Hamada, H., 2012. Two rotating cilia in the node cavity are sufficient to break left–right symmetry in the mouse embryo. *Nature Communications* **3**, 622.

Shiratori, H., and Hamada, H., 2006. The left-right axis in the mouse: from origin to morphology. *Development* **133**(11), 2095–2104.

Shiratori, H., Yashiro, K., Shen, M.M., and Hamada, H., 2006. Conserved regulation and role of

Pitx2 in situs-specific morphogenesis of visceral organs. *Development* **133**(15), 3015–3025.

Shlyueva, D., Stampfel, G., and Stark, A., 2014. Transcriptional enhancers: from properties to genome-wide predictions. *Nature reviews. Genetics* **15**(4), 272–86.

Shyu, A. Bin, Wilkinson, M.F., and van Hoof, A., 2008. Messenger RNA regulation: to translate or to degrade. *The EMBO journal* **27**(3), 471–481.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J.D., and Higgins, D.G., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**(1), 539.

Simon, E.M., Hevner, R.F., Pinter, J.D., Clegg, N.J., Delgado, M.R., Kinsman, S.L., Hahn, J.S., and Barkovich, A.J., 2002. The middle interhemispheric variant of holoprosencephaly. *American Journal of Neuroradiology* **23**(1), 151–155.

Singh, P., Schimenti, J.C., and Bolcun-Filas, E., 2014. A Mouse Geneticist's Practical Guide to CRISPR Applications. *Genetics* **199**(1), 1–15.

Sinha, T., Lin, L., Li, D., Davis, J., Evans, S., Wynshaw-Boris, A., and Wang, J., 2015. Mapping the dynamic expression of Wnt11 and the lineage contribution of Wnt11-expressing cells during early mouse development. *Developmental Biology* **398**(2), 177–92.

Smemo, S., Campos, L.C., Moskowitz, I.P., Krieger, J.E., Pereira, A.C., and Nobrega, M.A., 2012. Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Human Molecular Genetics* **21**(14), 3255–3263.

Solloway, M.J., Dudley, A.T., Bikoff, E.K., Lyons, K.M., Hogan, B.L.M., and Robertson, E.J., 1998. Mice lacking Bmp6 function. *Developmental Genetics* **22**(4), 321–339.

Solomon, B.D., Gropman, A., and Muenke, M., 1993. Holoprosencephaly Overview, GeneReviews.

Solomon, B.D., Lacbawan, F., Mercier, S., Clegg, N.J., Delgado, M.R., Rosenbaum, K., Dubourg, C., David, V., Olney, A.H., Wehner, L.-E., Hehr, U., Bale, S., Paulussen, A.D.C., Smeets, H.J., Hardisty, E., Tylki-Szymanska, A., Pronicka, E., Clemens, M., McPherson, E., Hennekam, R.C.M., Hahn, J.S., Stashinko, E., Levey, E., Wieczorek, D., Roeder, E., Schell-Apacik, C.C., Booth, C.W., Thomas, R.L., Kenwrick, S., Cummings, D.A.T., Bous, S.M., Keaton, A., Balog, J.Z., Hadley, D., Zhou, N., Long, R., Vélez, J.I., Pineda-Alvarez, D.E., Odent, S., Roessler, E., and Muenke, M., 2010a. Mutations in ZIC2 in human holoprosencephaly: description of a novel ZIC2 specific phenotype and comprehensive analysis of 157 individuals. *Journal of Medical Genetics* **47**(8), 513–524.

Solomon, B.D., Mercier, S., Vélez, J.I., Pineda-Alvarez, D.E., Wyllie, A., Zhou, N., Dubourg, C., David, V., Odent, S., Roessler, E., and Muenke, M., 2010b. Analysis of genotype-phenotype correlations in human holoprosencephaly. *American Journal of Medical Genetics, Part C: Seminars in Medical Genetics* **154**(1), 133–141.

Solomon, B.D., Pineda-Alvarez, D.E., Mercier, S., Raam, M.S., Odent, S., and Muenke, M., 2010c. Holoprosencephaly flashcards: A summary for the clinician. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics* **154C**(1), 3–7.

Solovyev, V., 2008. Statistical Approaches in Eukaryotic Gene Prediction. In: Handbook of Statistical Genetics. p. 1616.

Solovyev, V., and Salamov, A., 1997. The Gene-Finder computer tools for analysis of human and model organisms genome sequences. *Proceedings from the International Conference on Intelligent Systems for Molecular Biology* **5**, 294–302.

Solovyev, V. V., and Shahmuradov, I.A., 2003. PromH: Promoters identification using orthologous genomic sequences. *Nucleic Acids Research* **31**(13), 3540–3545.

Sotak, B.N., and Gleeson, J.G., 2012. Can't get there from here: cilia and hydrocephalus. *Nature medicine* **18**(12), 1742–3.

Spandidos, A., Wang, X., Wang, H., and Seed, B., 2009. PrimerBank: A resource of human and mouse PCR primer pairs for gene expression detection and quantification. *Nucleic Acids Research* **38**(SUPPL.1).

Spitz, F., and Furlong, E.E.M., 2012. Transcription factors: from enhancer binding to developmental control. *Nature Reviews Genetics* **13**(9), 613–626.

Spoelgen, R., Hammes, A., Anzenberger, U., Zechner, D., Andersen, O.M., Jerchow, B., and Willnow, T.E., 2005. LRP2/megalin is required for patterning of the ventral telencephalon. *Development* **132**(2), 405–414.

Stadhouders, R., Pas, S.D., Anber, J., Voermans, J., Mes, T.H.M., and Schutten, M., 2010. The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5' nuclease assay. *The Journal of Molecular Diagnostics* **12**(1), 109–17.

Storm, E.E., Garel, S., Borello, U., Hebert, J.M., Martinez, S., McConnell, S.K., Martin, G.R., and Rubenstein, J.L.R., 2006. Dose-dependent functions of Fgf8 in regulating telencephalic patterning centers. *Development* **133**(9), 1831–1844.

Struhl, K., 1999. Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* **98**(1), 1–4.

Suh, N., Baehner, L., Moltzahn, F., Melton, C., Shenoy, A., Chen, J., and Blelloch, R., 2010. MicroRNA Function Is Globally Suppressed in Mouse Oocytes and Early Embryos. *Current Biology* **20**(3), 271–277.

Sulik, K., Dehart, D.B., Iangaki, T., Carson, J.L., Vrablic, T., Gesteland, K., and Schoenwolf, G.C., 1994. Morphogenesis of the murine node and notochordal plate. *Developmental Dynamics* **201**(3), 260–278.

Sutherland, M.J., Wang, S., Quinn, M.E., Haaning, A.M., and Ware, S.M., 2013. Zic3 is required in the migrating primitive streak for node morphogenesis and left-right patterning. *Human Molecular Genetics* **22**(10), 1913–1923.

Tam, P.P., and Beddington, R.S., 1992. Establishment and organization of germ layers in the gastrulating mouse embryo. *Ciba Found Symp* **165**, 27–41.

Tam, P.P., and Steiner, K.A., 1999. Anterior patterning by synergistic activity of the early gastrula organizer and the anterior germ layer tissues of the mouse embryo. *Development* **126**(22), 5171–5179.

Tam, P.P.L., and Behringer, R.R., 1997. Mouse gastrulation: the formation of a mammalian body plan. *Mechanisms of Development* **68**(1–2), 3–25.

Tam, P.P.L., Loebel, D.A.F., and Tanaka, S.S., 2006. Building the mouse gastrula: signals, asymmetry and lineages. *Current opinion in genetics & development* **16**(4), 419–25.

Tam, P.P.L., Steiner, K.A., Zhou, S.X., and Quinlan, G.A., 1997. Lineage and functional analyses of the mouse organizer. *Cold Spring Harbor Symposia on Quantitative Biology* **62**, 135–144.

Tamm, C., Kadekar, S., Pijuan-Galitó, S., and Annerén, C., 2016. Fast and Efficient Transfection of Mouse Embryonic Stem Cells Using Non-Viral Reagents. *Stem cell reviews* **12**(5), 584–591.

Taylor, J.K., Levy, T., Suh, E.R., and Traber, P.G., 1997. Activation of enhancer elements by the homeobox gene Cdx2 is cell line specific. *Nucleic Acids Research* **25**(12), 2293–2300.

Tesar, P.J., Chenoweth, J.G., Brook, F.A., Davies, T.J., Evans, E.P., Mack, D.L., Gardner, R.L., and McKay, R.D.G., 2007. New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**(7150), 196–199.

Theil, T., Alvarez-Bolado, G., Walter, a, and Rüther, U., 1999. Gli3 is required for Emx gene expression during dorsal telencephalon development. *Development* **126**, 3561–3571.

Thomas, J., Morlé, L., Soulavie, F., Laurençon, A., Sagnol, S., and Durand, B., 2010. Transcriptional control of genes involved in ciliogenesis: a first step in making cilia. *Biology of the Cell* **102**(9), 499–513.

Thomsen, N., Ali, R.G., Ahmed, J.N., Arkell, R.M., and Voelkerding, K., 2012. High Resolution Melt Analysis (HRMA); a Viable Alternative to Agarose Gel Electrophoresis for Mouse Genotyping. *PLoS ONE* **7**(9), e45252.

Tian, B., Hu, J., Zhang, H., and Lutz, C.S., 2005. A large-scale analysis of mRNA polyadenylation of human and mouse genes. *Nucleic Acids Research* **33**(1), 201–212.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and De Laat, W., 2002. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Molecular Cell* **10**(6), 1453–1465.

Tombácz, D., Csabai, Z., Oláh, P., Balázs, Z., Likó, I., Zsigmond, L., Sharon, D., Snyder, M., and Boldogkői, Z., 2016. Full-length isoform sequencing reveals novel transcripts and substantial transcriptional overlaps in a herpesvirus. *PLoS ONE* **11**(9), e0162868.

Touriol, C., Morillon, A., Gensac, M.-C., Prats, H., and Prats, A.-C., 1999. Expression of Human Fibroblast Growth Factor 2 mRNA Is Post-transcriptionally Controlled by a Unique Destabilizing Element Present in the 3'-Untranslated Region between Alternative Polyadenylation Sites. *Journal of Biological Chemistry* **274**(30), 21402–21408.

Valencia-Sanchez, M.A., Liu, J., Hannon, G.J., and Parker, R., 2006. Control of translation and mRNA degradation by miRNAs and siRNAs. *Genes & Development* **20**(5), 515–524.

Vavouri, T., and Lehner, B., 2009. Conserved noncoding elements and the evolution of animal body plans. *BioEssays* **31**(7), 727–735.

Vavouri, T., McEwen, G.K., Woolfe, A., Gilks, W.R., and Elgar, G., 2006. Defining a genomic radius for long-range enhancer action: Duplicated conserved non-coding elements hold the key. *Trends in Genetics* **22**(1), 5–10.

Vavouri, T., Walter, K., Gilks, W.R., Lehner, B., and Elgar, G., 2007. Parallel evolution of conserved non-coding elements that target a common set of developmental regulatory genes from worms to humans. *Genome Biology* **8**(2), R15.

Vernimmen, D., De Gobbi, M., Sloane-Stanley, J.A., Wood, W.G., and Higgs, D.R., 2007. Long-

range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *The EMBO journal* **26**(8), 2041–51.

Vincent, S.D., Dunn, N.R., Hayashi, S., Norris, D.P., and Robertson, E.J., 2003. Cell fate decisions within the mouse organizer are governed by graded Nodal signals. *Genes & Development* **17**(13), 1646–1662.

Viotti, M., Niu, L., Shi, S.H., and Hadjantonakis, A.-K., 2012. Role of the gut endoderm in relaying left-right patterning in mice. *PLoS Biology* **10**(3).

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J.A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., Afzal, V., Ren, B., Rubin, E.M., and Pennacchio, L.A., 2009. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**(7231), 854–858.

Visel, A., Prabhakar, S., Akiyama, J.A., Shoukry, M., Lewis, K.D., Holt, A., Plajzer-Frick, I., Afzal, V., Rubin, E.M., and Pennacchio, L.A., 2008. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nature Genetics* **40**(2), 158–160.

Vlasova-St. Louis, I., and Bohjanen, P.R., 2011. Coordinate regulation of mRNA decay networks by GU-rich elements and CELF1. *Current Opinion in Genetics and Development* **21**(4), 444–451.

Vlasova, I.A., Tahoe, N.M., Fan, D., Larsson, O., Rattenbacher, B., SternJohn, J.R., Vasdewani, J., Karypis, G., Reilly, C.S., Bitterman, P.B., and Bohjanen, P.R., 2008. Conserved GU-Rich Elements Mediate mRNA Decay by Binding to CUG-Binding Protein 1. *Molecular Cell* **29**(2), 263–270.

Wall, R.J., 2001. Pronuclear microinjection. *Cloning and stem cells* **3**(4), 209–220.

Wang, W., Lin, C., Lu, D., Ning, Z., Cox, T., Melvin, D., Wang, X., Bradley, A., and Liu, P., 2008. Chromosomal transposition of PiggyBac in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the United States of America* **105**(27), 9290–9295.

Ware, S.M., Peng, J.L., Zhu, L., Fernbach, S.D., Colicos, S., Casey, B., Towbin, J. a, and Belmont, J.W., 2004. Identification and functional analysis of ZIC3 mutations in heterotaxy and related congenital heart defects. *American Journal of Human Genetics* **74**(1), 93–105.

Warr, N., Powles-Glover, N., Chappell, A., Robson, J., Norris, D., and Arkell, R.M., 2008. Zic2-associated holoprosencephaly is caused by a transient defect in the organizer region during gastrulation. *Human Molecular Genetics* **17**(19), 2986–96.

Wasserman, W.W., and Sandelin, A., 2004. Applied bioinformatics for the identification of regulatory elements. *Nature Reviews Genetics* **5**(4), 276–287.

Weill, L., Belloc, E., Bava, F.-A., and Méndez, R., 2012. Translational control by changes in poly(A) tail length: recycling mRNAs. *Nature Structural & Molecular Biology* **19**(6), 577–585.

Weinstein, D.C., Ruiz i Altaba, A., Chen, W.S., Hoodless, P., Prezioso, V.R., Jessell, T.M., and Darnell, J.E., 1994. The winged-helix transcription factor HNF-3 beta is required for notochord development in the mouse embryo. *Cell* **78**(4), 575–88.

Weiss, K., Kruszka, P., Guillen Sacoto, M.J., Addissie, Y.A., Hadley, D.W., Hadsall, C.K., Stokes, B., Hu, P., Roessler, E., Solomon, B., Wiggs, E., Thurm, A., Hufnagel, R.B., Zein, W.M., Hahn, J.S., Stashinko, E., Levey, E., Baldwin, D., Clegg, N.J., Delgado, M.R., and Muenke, M., 2017. In-depth investigations of adolescents and adults with holoprosencephaly identify unique characteristics. *Genetics in Medicine* **ePub**.

Wells, J.M., and Melton, D.A., 1999. Vertebrate Endoderm Development. *Annual Review of Cell and Developmental Biology* **15**(1), 393–410.

Wightman, B., Ha, I., and Ruvkun, G., 1993. Posttranscriptional regulation of the heterochronic gene lin-14 by lin- 4 mediates temporal pattern formation in C. elegans. *Cell* **75**(5), 855–862.

Wilkinson, D.G., 1992. Whole mount in situ hybridization of vertebrate embryos. In: In Situ Hybridization: A Practical Approach. IRL Press, Oxford, pp. 75–83.

Winnier, G., Blessing, M., Labosky, P.A., and Hogan, B.L., 1995. Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes & Development* **9**(17), 2105–16.

Winzi, M.K., Hyttel, P., Dale, J.K., and Serup, P., 2011. Isolation and characterization of node/notochord-like cells from mouse embryonic stem cells. *Stem cells and development* **20**(11), 1817–27.

Wood, H.B., and Episkopou, V., 1999. Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages. *Mechanisms of Development* **86**(1–2), 197–201.

Woolfe, A., Goodson, M., Goode, D.K., Snell, P., McEwen, G.K., Vavouri, T., Smith, S.F., North, P., Callaway, H., Kelly, K., Walter, K., Abnizova, I., Gilks, W.R., Edwards, Y.J.K., Cooke, J.E., and Elgar, G., 2005. Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biology* **3**(1).

Wu, S.C.-Y.C.-Y., Meir, Y.-J.J.Y.-J.J., Coates, C.J., Handler, A.M., Pelczar, P., Moisyadi, S., and Kaminski, J.M., 2006. piggyBac is a flexible and highly active transposon as compared to sleeping beauty, Tol2, and Mos1 in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **103**(41), 15008–15013.

Wu, X., and Brewer, G., 2012. The regulation of mRNA stability in mammalian cells: 2.0. *Gene* **500**(1), 10–21.

Xavier, G.M., Seppala, M., Barrell, W., Birjandi, A.A., Geoghegan, F., and Cobourne, M.T., 2016. Hedgehog receptor function during craniofacial development. *Developmental Biology* **415**(2), 198–215.

Xu, P., Vernooy, S.Y., Guo, M., and Hay, B.A., 2003. The Drosophila microRNA mir-14 suppresses cell death and is required for normal fat metabolism. *Current Biology* **13**(9), 790–795.

Yamamoto, M., Mine, N., Mochida, K., Sakai, Y., Saijoh, Y., Meno, C., and Hamada, H., 2003. Nodal signaling induces the midline barrier by activating Nodal expression in the lateral plate. *Development* **130**(9), 1795–1804.

Yamanaka, Y., Tamplin, O.J., Beckers, A., Gossler, A., and Rossant, J., 2007. Live imaging and genetic analysis of mouse notochord formation reveals regional morphogenetic mechanisms. *Developmental Cell* **13**(6), 884–896.

Yáñez-Cuna, J.O., Dinh, H.Q., Kvon, E.Z., Shlyueva, D., and Stark, A., 2012. Uncovering cis-regulatory sequence requirements for context-specific transcription factor binding. *Genome Research* **22**(10), 2018–2030.

Yang, A., Zhu, Z., Kapranov, P., McKeon, F., Church, G.M., Gingeras, T.R., and Struhl, K., 2006. Relationships between p63 binding, DNA sequence, transcription activity, and biological

function in human cells. *Molecular Cell* **24**(4), 593–602.

Yang, E., van Nimwegen, E., Zavolan, M., Rajewsky, N., Schroeder, M., Magnasco, M., and Darnell, J.E., 2003. Decay rates of human mRNAs: Correlation with functional characteristics and sequence attributes. *Genome Research* **13**(8), 1863–1872.

Yang, H., Wang, H., and Jaenisch, R., 2014. Generating genetically modified mice using CRISPR/Cas-mediated genome engineering. *Nature Protocols* **9**(8), 1956–1968.

Yavatkar, A.S., Lin, Y., Ross, J., Fann, Y., Brody, T., and Odenwald, W.F., 2008. Rapid detection and curation of conserved DNA via enhanced-BLAT and EvoPrinterHD analysis. *BMC Genomics* **9**, 106.

Ybot-Gonzalez, P., Gaston-Massuet, C., Girdler, G., Klingensmith, J., Arkell, R., Greene, N.D.E., and Copp, A.J., 2007. Neural plate morphogenesis during mouse neurulation is regulated by antagonism of Bmp signalling. *Development* **134**(17), 3203–3211.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., and Madden, T.L., 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**(1), 134.

Yee, S.P., and Rigby, P.W., 1993. The regulation of myogenin gene expression during the embryonic development of the mouse. *Genes & Development* **7**(7A), 1277–1289.

Yie, J., Merika, M., Munshi, N., Chen, G., and Thanos, D., 1999. The role of HMG I(Y) in the assembly and function of the IFN-beta enhanceosome. *EMBO Journal* **18**(11), 3074–3089.

Yoshiba, S., and Hamada, H., 2014. Roles of cilia, fluid flow, and Ca2+ signaling in breaking of left-right symmetry. *Trends in Genetics* **30**(1), 10–17.

Yoshiba, S., Shiratori, H., Kuo, I.Y., Kawasumi, A., Shinohara, K., Nonaka, S., Asai, Y., Sasaki, G., Belo, J.A., Sasaki, H., Nakai, J., Dworniczak, B., Ehrlich, B.E., Pennekamp, P., and Hamada, H., 2012. Cilia at the node of mouse embryos sense fluid flow for left-right determination via Pkd2. *Science* **338**(6104), 226–31.

Yoshioka, H., Meno, C., Koshiba, K., Sugihara, M., Itoh, H., Ishimaru, Y., Inoue, T., Ohuchi, H., Semina, E. V., Murray, J.C., Hamada, H., and Noji, S., 1998. Pitx2, a bicoid-type homeobox gene, is involved in a lefty-signaling pathway in determination of left-right asymmetry. *Cell* **94**(3), 299–305.

Young, T., Rowland, J.E., van de Ven, C., Bialecka, M., Novoa, A., Carapuco, M., van Nes, J., de Graaff, W., Duluc, I., Freund, J.-N., Beck, F., Mallo, M., and Deschamps, J., 2009. Cdx and Hox genes differentially regulate posterior axial growth in mammalian embryos. *Developmental Cell* **17**(4), 516–26.

Ysla, R.M., Wilson, G.M., and Brewer, G., 2008. Chapter 3. Assays of adenylate uridylate-rich element-mediated mRNA decay in cells. *Methods in enzymology* **449**, 47–71.

Yu, W., McDonnell, K., Taketo, M.M., and Bai, C.B., 2008. Wnt signaling determines ventral spinal cord cell fates in a time-dependent manner. *Development* **135**(22).

Yuan, H., Corbi, N., Basilico, C., and Dailey, L., 1995. Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes & Development* **9**(21), 2635–45.

Yuasa, S., Itabashi, Y., Koshimizu, U., Tanaka, T., Sugimura, K., Kinoshita, M., Hattori, F., Fukami, S., Shimazaki, T., Ogawa, S., Okano, H., and Fukuda, K., 2005. Transient inhibition

of BMP signaling by Noggin induces cardiomyocyte differentiation of mouse embryonic stem cells. *Nature Biotechnology* **23**(5), 607–11.

Yusa, K., Zhou, L., Li, M.A., Bradley, A., and Craig, N.L., 2011. A hyperactive piggyBac transposase for mammalian applications. *Proceedings of the National Academy of Sciences of the United States of America* **108**(4), 1531–6.

Zaret, K.S., and Carroll, J.S., 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes & Development* **25**(21), 2227–2241.

Zerucha, T., Stühmer, T., Hatch, G., Park, B.K., Long, Q., Yu, G., Gambarotta, A., Schultz, J.R., Rubenstein, J.L.R., and Ekker, M., 2000. A highly conserved enhancer in the Dlx5/Dlx6 intergenic region is the site of cross-regulatory interactions between Dlx genes in the embryonic forebrain. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **20**(2), 709–21.

Zetsche, B., Gootenberg, J.S., Abudayyeh, O.O., Slaymaker, I.M., Makarova, K.S., Essletzbichler, P., Volz, S.E., Joung, J., Oost, J. van der, Regev, A., Koonin, E.V., and Zhang, F., 2015. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163**(3), 759–771.

Zhang, H., and Bradley, A., 1996. Mice deficient for BMP2 are nonviable and have defects in amnion/chorion and cardiac development. *Development* **122**(10), 2977–86.

Zhang, L., Kasif, S., Cantor, C.R., and Broude, N.E., 2004. GC/AT-content spikes as genomic punctuation marks. *Proceedings of the National Academy of Sciences of the United States of America* **101**(48), 16855–16860.

Zhang, M., Bolfing, M.F., Knowles, H.J., Karnes, H., and Hackett, B.P., 2004. Foxj1 regulates asymmetric gene expression during left-right axis patterning in mice. *Biochemical and Biophysical Research Communications* **324**(4), 1413–1420.

Zhang Lab, and MIT, 2015. Optimized CRISPR Design [WWW Document]. URL http://crispr.mit.edu/ (accessed 3.29.17).

Zhao, W., Siegel, D., Biton, A., Tonqueze, O. Le, Zaitlen, N., Ahituv, N., and Erle, D.J., 2017. CRISPR-Cas9-mediated functional dissection of 3'-UTRs. *Nucleic Acids Research* **45**(18), 10800–10810.

Zhou, H.Y., Katsman, Y., Dhaliwal, N.K., Davidson, S., Macpherson, N.N., Sakthidevi, M., Collura, F., and Mitchell, J.A., 2014. A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. *Genes & Development* **28**(24), 2699–711.

Zhou, X., Sasaki, H., Lowe, L., Hogan, B.L.M., and Kuehn, M.R., 1993. Nodal is a novel TGF-β-like gene expressed in the mouse node during gastrulation. *Nature* **361**(6412), 543–547.

Zhu, L., Belmont, J.W., and Ware, S.M., 2006. Genetics of human heterotaxias. *European Journal of Human Genetics* **14**(1), 17–25.

Zhu, X., Ling, J., Zhang, L., Pi, W., Wu, M., and Tuan, D., 2007. A facilitated tracking and transcription mechanism of long-range enhancer function. *Nucleic Acids Research* **35**(16), 5532–5544.

Zimmerman, L.B., De Jesús-Escobar, J.M., and Harland, R.M., 1996. The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4. *Cell* **86**(4), 599–606.

Zubiaga, A.M., Belasco, J.G., and Greenberg, M.E., 1995. The nonamer UUAUUUAUU is the key

AU-rich sequence motif that mediates mRNA degradation. *Molecular and Cellular Biology* **15**(4), 2219–2230.